INVARIANCE PRINCIPLES IN FUNCTIONAL TIME SERIES ANALYSIS WITH APPLICATIONS

by

Gregory Rice

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

The University of Utah

May 2015

Copyright © Gregory Rice 2015

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Gregory Rice**

has been approved by the following supervisory committee members:

Lajos Horvath	, Chair	**3/9/2015** Date Approved
Davar Khoshnevisan	, Member	**3/10/2015** Date Approved
Stewart Ethier	, Member	**3/9/2015** Date Approved
Tom Alberts	, Member	**3/9/2015** Date Approved
David Kiefer	, Member	**3/9/2015** Date Approved

and by **Peter Trapa** , Chair/Dean of

the Department/College/School of **Mathematics**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

This dissertation aims to develop the theory and applications of functional time series analysis. Functional data analysis came into prominence in the 1990s when more sophisticated data collection and storage systems became prevalent, and many of the early developments focused on simple random samples of curves. However, a common source of functional data is when long, continuous records are broken into segments of smaller curves. An example of this is geologic and economic data that are presented as hourly or daily curves. In these instances, successive curves may exhibit dependencies which invalidate statistical procedures that assume a simple random sample.

The theory of functional time series analysis has grown tremendously in the last decade to provide methodology for such data, and researchers have focused primarily on adapting methods available in finite dimensional time series analysis to the function space setting. As a first problem, we consider an invariance principle for the partial sum process of stationary random functions. This theory is then applied to the problems of testing for stationarity of a functional time series and the one-way functional analysis of variance problem under dependence.

For my family.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to begin by thanking my family, in particular my fiancé Brittney, my mother and father Paula and Kent, my brothers Brad and Doug, and my grandparents Merle, Wanda, Robert, and Jo. Words cannot express how much you all mean to me. I would not have accomplished a thing if not for what I learned from each of you, like the value of hard work, and friendly competition, and I thank you all for that. Thank you also for your unconditional support and love.

I was just a high school football player going through the motions of school when I entered Mr. Ward's trigonometry class as a senior. It would be the last math class I would take in high school, but Mr. Ward left a lasting impression on me. He pushed me during class, realized I had some talent for the subject, and asked me to enter a district-wide Mathematics contest, in which I got second place. I doubt I ever would have even considered studying mathematics if not for his encouragement. Thank you, Mr. Ward.

Among the many faculty members at Oregon State University who influenced me during my undergraduate studies, I would like to give a special thanks to Professor Edward Waymire. Professor Waymire met me weekly during my senior year to discuss and work on problems in Real Analysis. I feel so fortunate to have had that time to hone the mathematical skills that I find the most useful in my present work, and I always look back fondly on our meetings and conversations. Thank you, Professor Waymire, for your time, patience, and advice.

I would like to especially thank several members of the Mathematics Department at the University of Utah. Thank you, Professor Davar Koshnevisan, for so many wonderful courses. I learned a lot about both mathematics, and how to present it, by watching you work. Thank you, Professor Yekaterina Ephstein, for your advice and support. Thank you, Dr. Nelson Beebe, for a number of conversations on typesetting that led to an improved presentation of this thesis. I would also like to thank the graduate program coordinators during my time as a graduate student, Sandy Hiskey and Paula Tooman. Both of you were far too nice us, and you each saved me too many times to count; thank you for that.

I have been fortunate enough to have been able to collaborate with a number of great statisticians and mathematicians during the last four years from whom I learned a great deal. Thank you Professors István Berkes, Marie Hušková, and Piotr Kokoszka, for many fruitful conversations about mathematics, your advice, and your support.

My final acknowledgement goes to my thesis advisor Professor Lajos Horváth. Lajos treats his students more like his family than anything else. I was exceptionally fortunate to have had a supervisor who shared his thoughts, emphasized learning by doing, and pushed his students to succeed as Lajos did. I grew tremendously as a statistician under his tutelage. It is hard to count the hours that we spent together conversing, drinking espressos, and engaging in mathematical pursuits. I learned so much more from you than what is written in the pages below, and for that I am very grateful.

CHAPTER 1

INTRODUCTION

In this chapter, we introduce functional data analysis and functional time series analysis. Some examples of functional time series data are discussed. We then provide an outline of the problems considered and their organization in this dissertation.

1.1 Functional data, functional time series, and examples

In classical statistics, it is typically assumed that the observations are elements of $\mathbb{R}^d$, and that the sample size N is larger than, or at least comparable in size to, d. Modern data, though, often exhibits such high dimensionality ($d >> N$) that classical statistical procedures are invalid, and this necessitates the development of new theory. One type of such data is continuous time phenomena that are observed at a high frequency. For example, the left panel of Figure 1.1 shows linearly interpolated measurements over time of the horizontal intensity of the Earth's magnetic field, a measure of the Earth's magnetic field strength, taken in 2001. At the weather station that provided this data, the horizontal intensity was measured every minute, giving a total of 1440 measurements per day. These measurements approximate up to a 1-minute resolution how the horizontal intensity is changing over time. It is thus natural, in this case, to view the 1440 daily measurements as a discrete sample of an underlying daily *curve* that represents the horizontal intensity throughout the day. Similarly, even though the stock price of a company is only recorded each time it is traded, which constitutes millions of irregularly spaced observations per day, if one were to check the price of Disney stock at google.com/finance, it will be displayed as a continuous curve such as in the right hand panel of Figure 1.1. It is quite common that high-dimensional and high-frequency data can be interpreted as curves or functions.

Viewing data in this way is the basis of functional data analysis (FDA). What is gained by trading finite dimensional observations, like the raw data in the above examples, for in principle infinite dimensional curves is two fold. Firstly, when the data are generated by an underlying continuous phenomena, a curve-based analysis is more appropriate since it takes advantage of this structure. Secondly, passing the data to a curve provides a more flexible framework for such data than multivariate techniques as it does not require that there be the same number of observations per curve, nor that the observations be obtained at equally spaced time intervals.

A common way in which functional data are obtained is by breaking long, continuous records into segments of shorter, for example hourly or daily, curves. The points at which to segment raw data into individual curves are often clear in the context of the data. For example, with magnetogram records, segmenting the raw data into daily curves is natural due to the effect of the Earth's rotation on the magnetic field. In these cases, successive curves may exhibit dependencies, and functional time series analysis, which combines concepts from FDA and classical time series analysis, provides a theory to model and utilize these dependencies. This dissertation aims to develop the theory of functional time series analysis and corresponding methodology.

1.2 Organization of the dissertation

This dissertation is organized into three remaining chapters. In Chapter 2, we consider an invariance principle for the partial sum process of stationary random functions that exhibit a Bernoulli shift structure. Chapter 3 develops an application of this result to testing for the stationarity of a functional time series that may be considered as an analog of the popular KPSS test. The dissertation concludes with Chapter 4 in which the one-way functional analysis of variance problem under serial dependence within each population in considered.

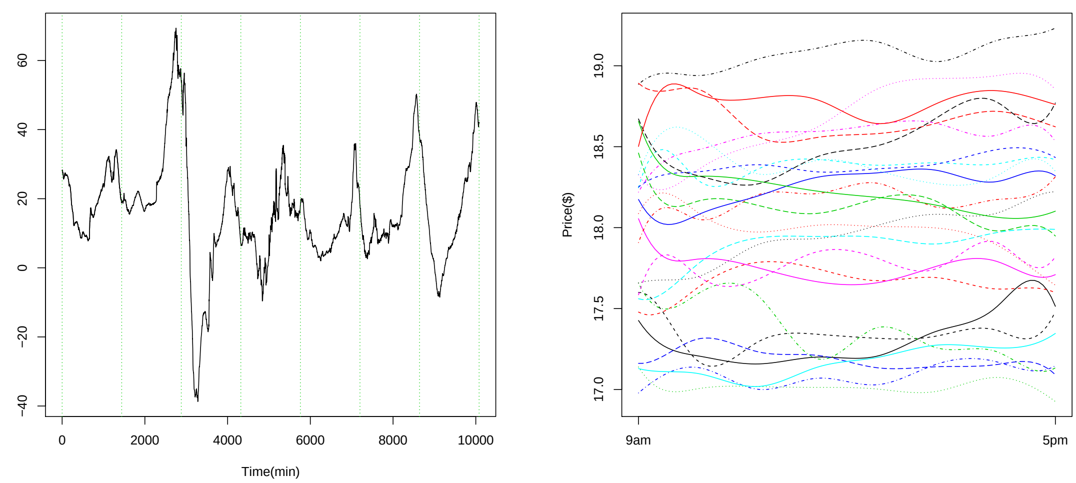

Figure 1.1. The graph on the left displays over 10,000 horizontal intensity measurements taken from March 1 to March 7, 2001. The green lines separate the data into daily functional observations. The graph on the right displays 25 functional observations derived from the intraday stock price of Disney; each curve represents one day of data.

CHAPTER 2

WEAK INVARIANCE PRINCIPLES FOR SUMS OF DEPENDENT RANDOM FUNCTIONS[1]

Motivated by problems in functional data analysis, in this chapter, we prove the weak convergence of normalized partial sums of dependent random functions exhibiting a Bernoulli shift structure.

2.1 Introduction

Functional data analysis in many cases requires central limit theorems and invariance principles for partial sums of random functions. The case of independent summands is much studied and well understood, but the theory for the dependent case is less complete. In this chapter, we study the important class of Bernoulli shift processes which are often used to model econometric and financial data. Let $\mathbf{X} = \{X_i(t)\}_{i=-\infty}^{\infty}$ be a sequence of random functions, square integrable on $[0, 1]$, and let $||\cdot||$ denote the $L^2[0, 1]$ norm. To lighten the notation, we use f for $f(t)$ when it does not cause confusion. Throughout this chapter, we assume that

$$\begin{aligned}&\mathbf{X} \text{ forms a sequence of Bernoulli shifts, i.e. } X_j(t) = g(\epsilon_j(t), \epsilon_{j-1}(t), ...) \text{ for some} \\ &\text{nonrandom measurable function } \ g : S^{\infty} \mapsto L^2 \ \text{ and iid random functions } \epsilon_j(t), \\ &-\infty < j < \infty, \text{ with values in a measurable space } \ S,\end{aligned} \tag{2.1}$$

$$\epsilon_j(t) = \epsilon_j(t, \omega) \ \text{ is jointly measurable in } \ (t, \omega) \ \ (-\infty < j < \infty), \tag{2.2}$$

$$EX_0(t) = 0 \text{ for all } t, \text{ and } E||X_0||^{2+\delta} < \infty \ \text{ for some } \ 0 < \delta < 1, \tag{2.3}$$

and

[1] The content of this chapter is based on joint research with István Berkes and Lajos Horváth.

the sequence $\{X_n\}_{n=-\infty}^{\infty}$ can be approximated by ℓ–dependent sequences (2.4)

$\{X_{n,\ell}\}_{n=-\infty}^{\infty}$ in the sense that

$$\sum_{\ell=1}^{\infty}(E||X_n - X_{n,\ell}||^{2+\delta})^{1/\kappa} < \infty \text{ for some } \kappa > 2+\delta,$$

where $X_{n,\ell}$ is defined by $X_{n,\ell} = g(\epsilon_n, \epsilon_{n-1}, ..., \epsilon_{n-\ell+1}, \boldsymbol{\epsilon}_{n,\ell}^*)$,

$\boldsymbol{\epsilon}_{n,\ell}^* = (\epsilon_{n,\ell,n-\ell}^*, \epsilon_{n,\ell,n-\ell-1}^*, \ldots)$, where the $\epsilon_{n,\ell,k}^*$'s are independent copies of ϵ_0,

independent of $\{\epsilon_i, -\infty < i < \infty\}$.

We note that assumption (2.1) implies that X_n is a stationary and ergodic sequence. Hörmann and Kokoszka (2010) call the processes satisfying (2.1)–(2.4) L^2 m–decomposable processes. The idea of approximating a stationary sequence with random variables which exhibit finite dependence first appeared in Ibragimov (1962) and is used frequently in the literature (cf. Billingsley (1968)). Aue et al.(2012) provide several examples when assumptions (2.1)–(2.4) hold which include autoregressive, moving average, and linear processes in Hilbert spaces. Also, the nonlinear functional ARCH(1) model (cf. Hörmann et al.(2012)) and bilinear models (cf. Hörmann and Kokoszka (2010)) satisfy (2.4).

We show in Section 2.2 (cf. Lemma 2.2.2) that the series in

$$C(t,s) = E[X_0(t)X_0(s)] + \sum_{\ell=1}^{\infty} E[X_0(t)X_\ell(s)] + \sum_{\ell=1}^{\infty} E[X_0(s)X_\ell(t)] \tag{2.5}$$

are convergent in L^2. The function $C(t,s)$ is positive definite, and therefore, there exist $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and orthonormal functions $\phi_i(t), 0 \leq t \leq 1$ satisfying

$$\lambda_i \phi_i(t) = \int C(t,s)\phi_i(s)ds, \quad 1 \leq i < \infty, \tag{2.6}$$

where $\int$ means $\int_0^1$. We define

$$\Gamma(x,t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x)\phi_i(t),$$

where W_i are independent and identically distributed Wiener processes (standard Brownian motions). Clearly, $\Gamma(x,t)$ is Gaussian. We show in Lemma 2.2.2 that $\sum_{\ell=1}^{\infty}\lambda_\ell < \infty$, and therefore,

$$\sup_{0 \leq x \leq 1} \int \Gamma^2(x,t)dt < \infty \quad \text{a.s.}$$

Theorem 2.1.1. *If assumptions (2.1)–(2.4) hold, then for every N we can define a Gaus-*

sian process $\Gamma_N(x,t)$ *such that*

$$\{\Gamma_N(x,t), 0 \le x, t \le 1\} \stackrel{\mathcal{D}}{=} \{\Gamma(x,t), 0 \le x, t \le 1\}$$

and

$$\sup_{0\le x\le 1} \int (S_N(x,t) - \Gamma_N(x,t))^2 dt = o_P(1),$$

where

$$S_N(x,t) = \frac{1}{N^{1/2}} \sum_{i=1}^{\lfloor Nx \rfloor} X_i(t).$$

The proof of Theorem 2.1.1 is given in Section 2.2. The proof is based on a maximal inequality which is given in Section 2.3 and is of interest in its own right. There is a wide literature on the central limit theorem for sums of random processes in abstract spaces. For limit theorems for sums of independent Banach space valued random variables, we refer to Ledoux and Talagrand (1991). For the central limit theory in the context of functional data analysis, we refer to the books of Bosq (2000) and Horváth and Kokoszka (2012). In the real valued case, the martingale approach to weak dependence was developed by Gordin (1969) and Philipp and Stout (1975), and by using such techniques, Merlevède (1996) and Dedecker and Merlevède (2003) obtained central limit theorems for a large class of dependent variables in Hilbert spaces. For some early influential results on invariance for sums of mixing variables in Banach spaces, we refer to Kuelbs and Philipp (1980), Dehling and Philipp (1982), and Dehling (1983). These papers provide very sharp results, but verifying mixing conditions is generally not easy and without additional continuity conditions, even autoregressive (1) processes may fail to be strong mixing (cf. Bradley (2007)). The weak dependence concept of Doukhan and Louhichi (1999) (cf. also Dedecker et al. (2007)) solves this difficulty, but so far, this concept has not been extended to variables in Hilbert spaces. Wu (2005, 2007) proved several limit theorems for one–dimensional stationary processes having a Bernoulli shift representation. Compared to classical mixing conditions, Wu's physical dependence conditions are easier to verify in concrete cases. Condition (2.3) cannot be directly compared to the approximating martingale conditions of Wu (2005, 2007). For extensions to the Hilbert space case, we refer to Hörmann and Kokoszka (2010).

2.2 Proof of theorem 2.1.1

The proof is based on three steps. We recall the definition of $X_{i,m}$ from (2.4). For every fixed m, the sequence $\{X_{i,m}\}$ is m–dependent. According to our first lemma, the sums of the X_i's can be approximated with the sums of m–dependent variables. The second step is the approximation of the infinite dimensional $X_{i,m}$'s with finite dimensional variables (Lemma 2.2.4). Then the result in Theorem 2.1.1 is established for finite dimensional m–dependent random functions (Lemma 2.2.6).

Lemma 2.2.1. *If (2.1)–(2.4) hold, then for all $x > 0$, we have*

$$\lim_{m\to\infty} \limsup_{N\to\infty} P\left\{ \max_{1\le k\le N} \frac{1}{\sqrt{N}} \left\| \sum_{i=1}^{k} (X_i - X_{i,m}) \right\| > x \right\} = 0. \tag{2.7}$$

Proof. The proof of this lemma requires the maximal inequality of Theorem 2.3.2. Section 2.3 is devoted to the proof of this result. Using Theorem 2.3.2, (2.7) is an immediate consequence of Markov's inequality. □

Define

$$C_m(t,s) = E[X_{0,m}(t)X_{0,m}(s)] + \sum_{i=1}^{m} E[X_{0,m}(t)X_{i,m}(s)] + \sum_{i=1}^{m} E[X_{0,m}(s)X_{i,m}(t)]. \tag{2.8}$$

We show in the following lemma that for every m, the function C_m is square–integrable. Hence there are $\lambda_{1,m} \ge \lambda_{2,m} \ge \cdots \ge 0$ and corresponding orthonormal functions $\phi_{i,m}, i = 1, 2, \ldots$ satisfying

$$\lambda_{i,m}\phi_{i,m}(t) = \int C_m(t,s)\phi_{i,m}(s)ds, \quad i = 1, 2, \ldots$$

Lemma 2.2.2. *If (2.1)–(2.4) hold, then we have*

$$\iint C^2(t,s)dtds < \infty, \tag{2.9}$$

$$\iint C_m^2(t,s)dtds < \infty \quad \textit{for all} \;\; m \ge 1, \tag{2.10}$$

$$\lim_{m\to\infty} \iint (C(t,s) - C_m(t,s))^2 dtds = 0. \tag{2.11}$$

$$\int C(t,t)dt = \sum_{k=1}^{\infty} \lambda_k < \infty, \tag{2.12}$$

$$\int C_m(t,t)dt = \sum_{k=1}^{\infty} \lambda_{k,m} < \infty \tag{2.13}$$

and

$$\lim_{m\to\infty} \int C_m(t,t)dt = \int C(t,t)dt. \tag{2.14}$$

Proof. Using the Cauchy-Schwarz inequality for expected values, we get

$$\iint (E[X_0(t)X_0(s)])^2 dtds \le \iint ((EX_0^2(t))^{1/2}(EX_0^2(s))^{1/2})^2 dtds = (E||X_0||^2)^2 < \infty.$$

Recalling that X_0 and $X_{i,i}$ are independent and both have 0 mean, we conclude first using the triangle inequality and then the Cauchy–Schwarz inequality for expected values that

$$\begin{aligned}
&\left\{\iint \left(\sum_{i=1}^{\infty} E[X_0(t)X_i(s)]\right)^2 dtds\right\}^{1/2} \qquad (2.15)\\
&= \left\{\iint \left(\sum_{i=1}^{\infty} E[X_0(t)(X_i(s) - X_{i,i}(s))]\right)^2 dtds\right\}^{1/2}\\
&\le \left(\iint \left(\sum_{i=1}^{\infty} E|X_0(t)(X_i(s) - X_{i,i}(s))|\right)^2 dtds\right)^{1/2}\\
&\le \sum_{i=1}^{\infty} \left(\int\int \{E|X_0(t)(X_i(s) - X_{i,i}(s))|\}^2 dtds\right)^{1/2}\\
&\le \sum_{i=1}^{\infty} \iint \left\{(EX_0^2(t))^{1/2}(E(X_i(s) - X_{i,i}(s))^2)^{1/2}\right\}^2 dtds\\
&= \int EX_0^2(t)dt \sum_{i=1}^{\infty} \int E(X_i(s) - X_{i,i}(s))^2 ds\\
&= E||X_0||^2 \sum_{i=1}^{\infty} E||X_0 - X_{0,i}||^2\\
&< \infty
\end{aligned}$$

on account of (2.4). This completes the proof of (2.9). □
ch

Since $EX_{0,m}(t)X_{0,m}(s) = EX_0(t)X_0(s)$, in order to establish (2.10), it is enough to show that

$$\iint \left\{\sum_{i=1}^{m} E[X_{0,m}(t)X_{i,m}(s)]\right\}^2 dtds < \infty.$$

It follows from the definition of $X_{i,m}$ that the vectors $(X_{0,m}, X_{i,m})$ and $(X_0, X_{i,m})$ have the same distribution for all $1 \le i \le m$. Also, $(X_{i,m}, X_{i,i})$ has the same distribution as $(X_0, X_{0,i}), 1 \le i \le m$. Hence following the arguments in (2.15), we get

$$\begin{aligned}
\left\{\iint \left\{\sum_{i=1}^{m} |EX_{0,m}(t)X_{i,m}(s)|\right\}^2 dtds\right\}^{1/2} &= \left\{\int\int \left\{\sum_{i=1}^{m} |EX_0(t)X_{i,m}(s)|\right\}^2 dtds\right\}^{1/2}\\
&\le E||X_0||^2 \sum_{i=1}^{m} \int E(X_{i,m}(s) - X_{i,i}(s))^2 ds
\end{aligned}$$

$$\leq E||X_0||^2 \sum_{i=1}^{\infty} E||X_0 - X_{0,i}||^2.$$
$$< \infty.$$

The proof of (2.10) is now complete. The arguments used above also prove (2.11). Repeating the previous arguments we have

$$\begin{aligned}
\int C(t,t)dt &\leq \int EX_0^2(t)dt + 2\sum_{i=1}^{\infty}\int |E[X_0(t)X_i(t)]|dt \\
&= \int EX_0^2(t)dt + 2\sum_{i=1}^{\infty}\int |E[X_0(t)(X_i(t) - X_{i,i}(t))]|dt \\
&= \int EX_0^2(t)dt + 2\sum_{i=1}^{\infty}\int (EX_0^2(t))^{1/2}(E[X_i(t) - X_{i,i}(t)]^2)^{1/2}dt \\
&\leq E||X_0||^2 + 2\sum_{i=1}^{\infty}\left(\int EX_0^2(t)dt\right)^{1/2}\left(\int E[X_i(t) - X_{i,i}(t)]^2dt\right)^{1/2} \\
&= E||X_0||^2 + 2(E||X_0||^2)^{1/2}\sum_{i=1}^{\infty}(E||X_0 - X_{0,i}||^2)^{1/2} \\
&< \infty.
\end{aligned}$$

Observing that

$$\int C(t,t)dt = \sum_{i=1}^{\infty}\lambda_i \int \phi_i^2(t)dt = \sum_{i=1}^{\infty}\lambda_i,$$

the proof of (2.12) is complete. The same arguments can be used to establish (2.13). The relation in (2.14) can be established along the lines of the proof of (2.11). □

By the Karhunen–Loéve expansion, we have that

$$X_{i,m}(t) = \sum_{\ell=1}^{\infty}\langle X_{i,m}, \phi_{\ell,m}\rangle \phi_{\ell,m}(t). \tag{2.16}$$

Define

$$X_{i,m,K}(t) = \sum_{\ell=1}^{K}\langle X_{i,m}, \phi_{\ell,m}\rangle \phi_{\ell,m}(t) \tag{2.17}$$

to be the partial sums of the series in (2.16), and

$$\bar{X}_{i,m,K}(t) = X_{i,m}(t) - X_{i,m,K}(t) = \sum_{\ell=K+1}^{\infty}\langle X_{i,m}, \phi_{\ell,m}\rangle \phi_{\ell,m}(t). \tag{2.18}$$

Lemma 2.2.3. *If $\{Z_i\}_{i=1}^N$ are independent L^2 valued random variables such that*

$$EZ_1(t) = 0 \ \ and \ \ E||Z_1||^2 < \infty, \tag{2.19}$$

then for all $x > 0$, we have that

$$P\left\{\max_{1 \le k \le N} \left|\left|\sum_{i=1}^{k} Z_i\right|\right|^2 > x\right\} \le \frac{1}{x} E \left|\left|\sum_{i=1}^{N} Z_i\right|\right|^2. \tag{2.20}$$

Proof. Let $\mathcal{F}_k$ be the sigma algebra generated by the random variables $\{Z_j\}_{j=1}^k$. By assumption (2.19) and the independence of the $Z_i's$, we have that

$$E\left(\left|\left|\sum_{i=1}^{k+1} Z_i\right|\right|^2 \middle| \mathcal{F}_k\right) = \left|\left|\sum_{i=1}^{k} Z_i\right|\right|^2 + E||Z_{k+1}||^2 \ge \left|\left|\sum_{i=1}^{k} Z_i\right|\right|^2.$$

Therefore, $\left\{\left|\left|\sum_{i=1}^k Z_i\right|\right|^2\right\}_{k=1}^{\infty}$ is a non-negative submartingale with respect to the filtration $\{\mathcal{F}_k\}_{k=1}^{\infty}$. If we define

$$A = \left\{\omega : \max_{1 \le k \le N} \left|\left|\sum_{i=1}^{k} Z_i\right|\right|^2 > x\right\},$$

then it follows from Doob's maximal inequality (Chow and Teicher, 1988 p. 247) that

$$\begin{aligned} xP\left\{\max_{1 \le k \le N} \left|\left|\sum_{i=1}^{k} Z_i\right|\right|^2 > x\right\} &\le E\left(\left|\left|\sum_{i=1}^{N} Z_i\right|\right|^2 I_A\right) \\ &\le E\left|\left|\sum_{i=1}^{N} Z_i\right|\right|^2, \end{aligned}$$

which completes the proof. □

Lemma 2.2.4. *If (2.1)–(2.4) hold, then for all $x > 0$,*

$$\lim_{K \to \infty} \limsup_{N \to \infty} P\left\{\max_{1 \le k \le N} \left|\left|\frac{1}{\sqrt{N}} \sum_{i=1}^{k} \bar{X}_{i,m,K}\right|\right| > x\right\} = 0. \tag{2.21}$$

Proof. Define $Q_k(j) = \{i : 1 \le i \le k, i = j(mod \ \ m)\}$ for $j = 0, 1, ..., m-1$, and all positive integers k. It is then clear that

$$\sum_{i=1}^{k} \bar{X}_{i,m,K} = \sum_{j=0}^{m-1} \sum_{i \in Q_k(j)} \bar{X}_{i,m,K}.$$

We thus obtain by the triangle inequality that

$$P\left\{\max_{1\le k\le N}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{k}\bar{X}_{i,m,K}\right|\right|>x\right\}\le P\left\{\sum_{j=0}^{m-1}\max_{1\le k\le N}\left|\left|\frac{1}{\sqrt{N}}\sum_{i\in Q_k(j)}\bar{X}_{i,m,K}\right|\right|>x\right\}.$$

It is therefore sufficient to show that for each fixed j,

$$\lim_{K\to\infty}\limsup_{N\to\infty}P\left\{\max_{1\le k\le N}\left|\left|\frac{1}{\sqrt{N}}\sum_{i\in Q_k(j)}\bar{X}_{i,m,K}\right|\right|>x\right\}=0.$$

By the definition of $Q_k(j)$, $\{\bar{X}_{i,m,K}\}_{i\in Q_k(j)}$ is an iid sequence of random variables. So, by applications of Lemma 2.2.3 and the assumption (2.3), we have that

$$\begin{aligned}P\left\{\max_{1\le k\le N}\left|\left|\frac{1}{\sqrt{N}}\sum_{i\in Q_k(j)}\bar{X}_{i,m,K}\right|\right|^2>x\right\}&\le\frac{1}{x}E\left|\left|\frac{1}{\sqrt{N}}\sum_{i\in Q_N(j)}\bar{X}_{i,m,K}\right|\right|^2\\&\le\frac{1}{x}E||\bar{X}_{0,m,K}^2||\\&=\frac{1}{x}\sum_{\ell=K+1}^{\infty}\lambda_{\ell,m}.\end{aligned}\tag{2.22}$$

Since the right-hand side of (2.22) tends to zero as K tends to infinity independently of N, (2.21) follows. □

Clearly, with $k=\lfloor Nx\rfloor$ we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{k}X_{i,m,K}(t)=\sum_{j=1}^{K}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{j,m}\rangle\right)\phi_{j,m}(t).\tag{2.23}$$

Lemma 2.2.5. *If (2.1)–(2.4) hold, then the K dimensional random process*

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{1,m}\rangle,...,\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{K,m}\rangle\right)$$

converges, as $N\to\infty$, in $\mathcal{D}[0,1]$ to

$$\left(\lambda_{1,m}^{1/2}W_1(x),...,\lambda_{K,m}^{1/2}W_K(x)\right),\tag{2.24}$$

where $\{W_i\}_{i=1}^{K}$ are independent, identically distributed Wiener processes.

Proof. A similar procedure as in Lemma 2.2.4 shows that for each j, $\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{j,m}\rangle$ can be written as a sum of sums of independent and identically distributed random variables, and thus, by Billingsley (1968), it is tight. This implies that the K dimensional process

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{1,m}\rangle,...,\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{K,m}\rangle\right)$$

is tight, since it is tight in each coordinate. Furthermore, the Cramér-Wold device and the central limit theorem for m–dependent random variables (cf. DasGupta (2008) p. 119) shows that the finite dimensional distributions of the vector process converge to the finite dimensional distributions of the process in (2.24). The lemma follows. □

In light of the Skorkohod–Dudley–Wichura theorem (cf. Shorack and Wellner (1986), p. 47), we may reformulate Lemma 2.2.5 as follows.

Corollary 2.2.1. *If (2.1)–(2.4) hold, then for each positive integer N, there exists K independent, identically distributed Wiener processes $\{W_{i,N}\}_{i=1}^{K}$ such that for each j,*

$$\sup_{0\le x\le 1}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{j,m}\rangle-\lambda_{j,m}^{1/2}W_{j,N}(x)\right|\xrightarrow{P}0,$$

as $N\to\infty$.

Lemma 2.2.6. *If (2.1)–(2.4) hold, then for $\{W_{i,N}\}_{i=1}^{K}$ defined in Corollary 2.2.1, we have that*

$$\sup_{0\le x\le 1}\int\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}X_{i,m,K}(t)-\sum_{\ell=1}^{K}\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\phi_{\ell,m}(t)\right)^2dt\xrightarrow{P}0, \tag{2.25}$$

as $N\to\infty$.

Proof. By using (2.23), we get that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}X_{i,m,K}(t)-\sum_{\ell=1}^{K}\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\phi_{\ell,m}(t)$$
$$=\sum_{\ell=1}^{K}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{\ell,m}\rangle-\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\right)\phi_{\ell,m}(t).$$

The substitution of this into the expression in (2.25) along with a simple calculation shows that

$$\sup_{0\le x\le 1}\int\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}X_{i,m,K}(t)-\sum_{\ell=1}^{K}\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\phi_{\ell,m}(t)\Big)^2dt$$
$$=\sup_{0\le x\le 1}\sum_{\ell=1}^{K}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{\ell,m}\rangle-\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\Big)^2$$
$$\le\sum_{\ell=1}^{K}\sup_{0\le x\le 1}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}\langle X_{i,m},\phi_{\ell,m}\rangle-\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\Big)^2\xrightarrow{P}0,$$

as $N\to\infty$, by Corollary 2.2.1. □

Lemma 2.2.7. *If (2.1)–(2.4) hold,*

$$\sup_{0\le x\le 1}\int\left(\sum_{\ell=K+1}^{\infty}\lambda_{\ell,m}^{1/2}W_\ell(x)\phi_{\ell,m}(t)\right)^2dt\xrightarrow{P}0,\tag{2.26}$$

as $K\to\infty$, where $W_1,W_2,\ldots$ are independent and identically distributed Wiener processes.

Proof. Since the functions $\{\phi_{\ell,m}\}_{\ell=1}^{\infty}$ are orthonormal, we have that

$$E\sup_{0\le x\le 1}\int\left(\sum_{\ell=K+1}^{\infty}\lambda_{\ell,m}^{1/2}W_\ell(x)\phi_{\ell,m}(t)\right)^2dt=E\sup_{0\le x\le 1}\sum_{\ell=K+1}^{\infty}\lambda_{\ell,m}W_\ell^2(x)$$
$$\le\sum_{\ell=K+1}^{\infty}\lambda_{\ell,m}E\sup_{0\le x\le 1}W_\ell^2(x)\longrightarrow 0,$$

as $K\to\infty$. Therefore, (2.26) follows from the Markov inequality. □

Lemma 2.2.8. *If (2.1)–(2.4) hold, then for each N, we can define independent identically distributed Wiener processes $\{W_{i,N}\}_{i=1}^{K}$ such that*

$$\sup_{0\le x\le 1}\int\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nx\rfloor}X_{i,m}(t)-\sum_{\ell=1}^{\infty}\lambda_{\ell,m}^{1/2}W_{\ell,N}(x)\phi_{\ell,m}(t)\right)^2dt\xrightarrow{P}0,$$

as $N\to\infty$.

Proof. It follows from Lemmas 2.2.4-2.2.7. □

Since the distribution of $W_{\ell,N},1\le\ell<\infty$ does not depend on N, it is enough to consider the asymptotics for $\sum_{\ell=1}^{\infty}\lambda_{\ell,m}^{1/2}W_\ell(x)\phi_{\ell,m}(t)$, where W_ℓ are independent Wiener processes.

Lemma 2.2.9. *If (2.1)–(2.4) hold, then for each m, we can define independent and identically distributed Wiener processes $\bar W_{\ell,m}(x),1\le\ell<\infty$ such that*

$$\sup_{0\le x\le 1}\int\left(\sum_{\ell=1}^{\infty}\lambda_{\ell,m}^{1/2}W_\ell(x)\phi_{\ell,m}(t)-\sum_{\ell=1}^{\infty}\lambda_\ell^{1/2}\bar{W}_{\ell,m}(x)\phi_\ell(t)\right)^2dt\xrightarrow{P}0, \tag{2.27}$$

as $m\to\infty$.

Proof. Let

$$\Delta_m(x,t)=\sum_{\ell=1}^{\infty}\lambda_{\ell,m}^{1/2}W_\ell(x)\phi_{\ell,m}(t).$$

Let M be a positive integer and define $x_i=i/M, 0\le i\le M$. It is easy to see that

$$\begin{aligned}
&E\max_{0\le i<M}\sup_{0\le h\le 1/M}\int(\Delta_m(x_i+h,t)-\Delta_m(x_i,t))^2dt\\
&\quad\le\sum_{\ell=1}^{\infty}\lambda_{\ell,m}E\left\{\max_{0\le i<M}\sup_{0\le h\le 1/M}(W_\ell(x_i+h)-W_\ell(x_i))^2\right\}\\
&\quad=E\left\{\max_{0\le i<M}\sup_{0\le h\le 1/M}(W_1(x_i+h)-W_1(x_i))^2\right\}\sum_{\ell=1}^{\infty}\lambda_{\ell,m}.
\end{aligned}$$

Using Lemma 2.2.2, we get that

$$\sum_{\ell=1}^{\infty}\lambda_{\ell,m}=\int E\Delta_m^2(1,t)dt=\int C_m(t,t)dt\to\int C(t,t)dt=\sum_{\ell=1}^{\infty}\lambda_\ell.$$

So by the modulus of continuity of the Wiener process (cf. Garsia (1970)), we get that

$$\lim_{M\to\infty}\limsup_{m\to\infty}E\max_{0\le i<M}\sup_{0\le h\le 1/M}\int(\Delta_m(x_i+h,t)-\Delta_m(x_i,t))^2dt=0. \tag{2.28}$$

By the Karhunen-Loéve expansion, we can also write Δ_m as

$$\Delta_m(x,t)=\sum_{\ell=1}^{\infty}\langle\Delta_m(x,\cdot),\phi_\ell\rangle\phi_\ell(t)$$

and

$$E\int\Delta_m^2(x,t)dt=\sum_{\ell=1}^{\infty}E(\langle\Delta_m(x,\cdot),\phi_\ell\rangle)^2.$$

So by Lemma 2.2.2, we have

$$\sum_{\ell=1}^{\infty}E(\langle\Delta_m(x,\cdot),\phi_\ell\rangle)^2\to x\sum_{\ell=1}^{\infty}\lambda_\ell.$$

Also, for any positive integer ℓ,

$$E(\langle\Delta_m(x,\cdot),\phi_\ell\rangle)^2=\iint C_m(t,s)\phi_\ell(t)\phi_\ell(s)dtds\to\iint C(t,s)\phi_\ell(t)\phi_\ell(s)dtds=\lambda_\ell,$$

as $m\to\infty$. Hence for every $z>0$, we have

$$\limsup_{K\to\infty}\limsup_{m\to\infty} P\left\{\int\left(\sum_{\ell=K+1}^{\infty}\langle\Delta_m(x,\cdot),\phi_\ell\rangle\phi_\ell(t)\right)^2 dt > z\right\} = 0. \tag{2.29}$$

The joint distribution of $\langle\Delta(x_i,\cdot),\phi_\ell\rangle, 1 \le i \le M, 1 \le \ell \le K$ is multivariate normal with zero mean. Hence they converge jointly to a multivariate normal distribution. To show their joint convergence in distribution, we need to show the convergence of the covariance matrix. Using again Lemma 2.2.2, we get that

$$\begin{aligned} E\langle\Delta(x_i,\cdot),\phi_\ell\rangle\langle\Delta(x_j,\cdot),\phi_k\rangle &= \min(x_i,x_j)\iint C_m(t,s)\phi_\ell(t)\phi_k(s)dtds \\ \to \min(x_i,x_j)\iint C(t,s)\phi_\ell(t)\phi_k(s)dtds &= \min(x_i,x_j)\lambda_\ell I\{k=\ell\}. \end{aligned}$$

Due to this covariance structure and the Skorkohod–Dudley–Wichura theorem (cf. Shorack and Wellner (1986), p. 47), we can find independent Wiener processes $\bar{W}_{\ell,m}(x), 1 \le \ell < \infty$ such that

$$\max_{1\le i\le M}\max_{1\le \ell\le K}|\langle\Delta(x_i,\cdot),\phi_\ell\rangle - \lambda_\ell^{1/2}\bar{W}_{\ell,m}(x_i)| = o_P(1), \quad \text{as } m\to\infty.$$

Clearly, for all $0 \le x \le 1$

$$E\int\left(\sum_{\ell=K+1}^{\infty}\lambda_\ell^{1/2}\bar{W}_{\ell,m}(x)\phi_\ell(t)\right)^2 dt = x\sum_{\ell=K+1}^{\infty}\lambda_\ell \to 0, \quad \text{as } m\to\infty,$$

and therefore similarly to (2.29)

$$\limsup_{K\to\infty}\limsup_{m\to\infty} P\left\{\int\left(\sum_{\ell=K+1}^{\infty}\lambda_\ell^{1/2}\bar{W}_{\ell,m}(x)\phi_\ell(t)\right)^2 dt > z\right\} = 0$$

for all $z > 0$. Similarly to (2.28) one can show that

$$\begin{aligned} &E\max_{0\le i<M}\sup_{0\le h\le 1/M}\int\left(\sum_{\ell=1}^{\infty}(\bar{W}_{\ell,m}(x_i+h)-\bar{W}_{\ell,m}(x_i))\phi_\ell(t)\right)^2 dt \\ &\le E\left\{\max_{0\le i<M}\sup_{0\le h\le 1/M}(W(x_i+h)-W(x_i))^2\right\}\sum_{\ell=1}^{\infty}\lambda_\ell \to 0, \quad \text{as } M\to\infty, \end{aligned}$$

where W is a Wiener process. This also completes the proof of Lemma 2.2.9. □

Proof of Theorem 2.1.1 First we approximate $S_N(x,t)$ with m-dependent processes (Lemma 2.2.1). The second step of the proof is the approximation of the sums of m-dependent processes with a Gaussian process with covariance function $\min(x,x')C_m(t,s)$, where C_m is defined in (2.8) (Lemma 2.2.8)). The last step of the proof is the convergence of Gaussian processes with covariance functions $\min(x, x')C_m(t, s)$ to a Gaussian process with covariance function $\min(x, x')C(t, s)$ (Lemma 2.2.9).

2.3 Some moment and maximal inequalities

In this section, we give the proof of the maximal inequality in Lemma 2.2.1 which is a crucial ingredient of the proof of Theorem 2.1.1. Actually, we will prove below some moment and maximal inequalities for partial sums of function valued Bernoulli shift sequences which have their own interest and can be used in various related problems.

Our first lemma is a Hilbert space version of Doob's (1953 p. 226) inequality.

Lemma 2.3.1. *If Z_1 and Z_2 are independent mean zero Hilbert space valued random variables, and if $0 < \delta \le 1$, then*

$$E||Z_1 + Z_2||^{2+\delta} \le E||Z_1||^{2+\delta} + E||Z_2||^{2+\delta} + E||Z_1||^2 (E||Z_2||^2)^{\delta/2} + E||Z_2||^2 (E||Z_1||^2)^{\delta/2}.$$

Proof. Since $0 < \delta \le 1$, for any $A, B \ge 0$, we have that $(A + B)^\delta \le A^\delta + B^\delta$ (cf. Hardy et al.(1969, p. 32)). An application of this inequality along with Minkowski's inequality gives that

$$||Z_1 + Z_2||^\delta \le (||Z_1|| + ||Z_2||)^\delta \le ||Z_1||^\delta + ||Z_2||^\delta.$$

We also have by Hölders inequality that

$$E||Z_1||^\delta \le (E||Z_1||^2)^{\delta/2}.$$

This yields that

$$\begin{aligned}
E||Z_1 + Z_2||^{2+\delta} &= E||Z_1 + Z_2||^2 ||Z_1 + Z_2||^\delta \\
&\le E||Z_1 + Z_2||^2 (||Z_1||^\delta + ||Z_2||^\delta) \\
&= E[||Z_1||^2 + ||Z_2||^2 + 2\langle Z_1, Z_2\rangle](||Z_1||^\delta + ||Z_2||^\delta) \\
&= E||Z_1||^{2+\delta} + E||Z_2||^{2+\delta} + E||Z_1||^2 E||Z_2||^\delta + E||Z_2||^2 E||Z_1||^\delta \\
&\le E||Z_1||^{2+\delta} + E||Z_2||^{2+\delta} + E||Z_1||^2 (E||Z_2||^2)^{\delta/2} + E||Z_2||^2 (E||Z_1||^2)^{\delta/2},
\end{aligned}$$

which proves the lemma. □

Remark 2.3.1. If Z_1 and Z_2 are independent and identically distributed, then the result of Lemma 2.3.1 can be written as

$$E||Z_1 + Z_2||^{2+\delta} \le 2E||Z_1||^{2+\delta} + 2(E||Z_1||^2)^{1+\delta/2}.$$

Let

$$I(r) = \sum_{\ell=1}^{\infty} (E||X_0 - X_{0,\ell}||^r)^{1/r}. \tag{2.30}$$

We note that by (2.4), $I(r) < \infty$ for all $2 \le r \le 2+\delta$.

Lemma 2.3.2. *If (2.1)–(2.4) hold, then we have*

$$E\left|\left|\sum_{i=1}^{n}(X_i - X_{i,m})\right|\right|^2 \le nA,$$

where

$$A = \int E(X_0 - X_{0,m})^2(t)dt + 2^{5/2}\left(\int E(X_0 - X_{0,m})^2(t)dt\right)^{1/2} I(2). \tag{2.31}$$

Proof. Let $Y_i = X_i - X_{i,m}$. By Fubini's theorem and the fact that the random variables are identically distributed, we conclude

$$\begin{aligned} E\left|\left|\sum_{i=1}^{n} Y_i\right|\right|^2 &= \int E\left(\sum_{i=1}^{n} Y_i(t)\right)^2 dt \qquad (2.32)\\ &= n\int EY_0^2(t)dt + 2\int \sum_{i=1}^{n-1}(n-i)E[Y_0(t)Y_i(t)]dt \\ &\le n\int EY_0^2(t)dt + 2n\sum_{i=1}^{n-1}\int |E[Y_0(t)Y_i(t)]|dt \\ &\le n\int EY_0^2(t)dt + 2n\sum_{i=1}^{\infty}\int |E[Y_0(t)Y_i(t)]|dt. \end{aligned}$$

We recall $X_{i,i}$ from (2.4). Under this definition, the random variables Y_0 and $X_{i,i}$ are independent for all $i \ge 1$. Let $Z_i = X_{i,m}$, if $i > m$ and $Z_i = g(\epsilon_i, \ldots, \epsilon_1, \boldsymbol{\delta}_i)$, if $1 \le i \le m$, where $\boldsymbol{\delta}_i = (\delta_{i,0}, \delta_{i,-1}, \ldots)$ and $\delta_{i,j}$ are iid copies of ϵ_0, independent of the ϵ_ℓ's and $\boldsymbol{\epsilon}_{k,\ell}$'s. Clearly, Z_i and Y_0 are independent and thus with $Y_{i,i} = X_{i,i} - Z_i$ we have

$$E[Y_0(t)Y_i(t)] = E[Y_0(t)(Y_i(t) - Y_{i,i}(t))].$$

Furthermore, by first applying the Cauchy-Schwarz inequality for expected values and then by the Cauchy–Schwarz inequality for functions in L^2, we get that

$$\begin{aligned} \int |E[Y_0(t)(Y_i(t) - Y_{i,i}(t))]|dt &\le \int \left(EY_0^2(t)\right)^{1/2}\left(E\left[Y_i(t) - Y_{i,i}(t)\right]^2\right)^{1/2} dt \\ &\le \left(\int EY_0^2(t)dt\right)^{1/2}\left(\int E\left[Y_i(t) - Y_{i,i}(t)\right]^2 dt\right)^{1/2}. \end{aligned}$$

Also,

$$\int E\left[Y_i(t)-Y_{i,i}(t)\right]^2 dt \le 2\left(\int E\left[X_i(t)-X_{i,i}(t)\right]^2 dt+\int E\left[X_{i,m}(t)-Z_i(t)\right]^2 dt\right)$$

The substitution of this expression into (2.32) gives that

$$\begin{aligned}E\left|\left|\sum_{i=1}^{n} Y_i\right|\right|^2 &\le n\int EY_0^2(t)dt+2^{3/2}n\sum_{i=1}^{\infty}\left(\int EY_0^2(t)dt\right)^{1/2}\times\\ &\qquad\left\{\left(\int E\left[X_i(t)-X_{i,i}(t)\right]^2 dt\right)^{1/2}+\left(\int E\left[X_{i,m}(t)-Z_i(t)\right]^2 dt\right)^{1/2}\right\}\\ &\le n\left[\int EY_0^2(t)dt+2^{5/2}\left(\int EY_0^2(t)dt\right)^{1/2} I(2)\right],\end{aligned}$$

which completes the proof. □

Theorem 2.3.1. *If (2.1)–(2.4) hold, then for all* $N \ge 1$

$$E\left|\left|\sum_{i=1}^{N}(X_i-X_{i,m})\right|\right|^{2+\delta} \le N^{1+\delta/2}B,$$

where

$$\begin{aligned}B = E||X_0-X_{0,m}||^{2+\delta}&+c_\delta^{2+\delta}[A^{1+\delta/2}+J_m^{2+\delta}+J_mA^{(1+\delta)/2}+A^{(1+\delta/2)\delta}J_m^2]\\ &+(c_\delta J_m^2)^{1/(1-\delta)}\end{aligned} \tag{2.33}$$

with A *defined in (2.31),*

$$c_\delta = 36\left(1-\frac{1}{2^{\delta/2}}\right)^{-1} \tag{2.34}$$

and

$$J_m = 2(E||X_0-X_{0,m}||^{2+\delta})^{(\kappa-2-\delta)/(\kappa(2+\delta))}\sum_{\ell=1}^{\infty}(E||X_0-X_{0,\ell}||^{2+\delta})^{1/\kappa}.$$

Proof. We prove Theorem 2.3.1 using mathematical induction. By the definition of B, the inequality is obvious when $N=1$. Assume that it holds for all k which are less than or equal to $N-1$. We assume that N is even, i.e. $N=2n$. The case when N is odd can be done in the same way with minor modifications. Let $Y_i = X_i - X_{i,m}$. For all i satisfying $n+1 \le i \le 2n$, we define

$$X_{i,n}^* = g(\epsilon_i, \epsilon_{i-1}, ..., \epsilon_{n+1}, \epsilon_n^*, \epsilon_{n-1}^*, ...)$$

where the ϵ_j^*'s denote iid copies of ϵ_0, independent of $\{\epsilon_i, -\infty < i < \infty\}$ and $\{\boldsymbol{\epsilon}_{k,\ell}^*, -\infty < k, \ell < \infty\}$. We define $Z_{i,n} = X_{i,m}$, if $m+n+1 \le i \le 2n$ and

$$Z_{i,n} = g(\epsilon_i, \ldots, \epsilon_{n+1}, \epsilon_n^*, \ldots \epsilon_{i-m+1}^*, \boldsymbol{\delta}_i) \quad \text{with} \quad \boldsymbol{\delta}_i = (\delta_{i,n}, \delta_{i,n-1}, \ldots),$$

if $n+1 \le i \le n+m$, where the $\delta_{k,\ell}$'s are iid copies of ϵ_0, independent of the ϵ_k's and $\boldsymbol{\epsilon}_{k,\ell}^*$'s. Let $Y_{i,n}^* = X_{i,n}^* - Z_{i,n}$, if $n+1 \le i \le 2n$. Under this definition, the sequences $\{Y_i, 1 \le i \le n\}$ and $\{Y_{i,n}^*, n+1 \le i \le 2n\}$ are independent and have the same distribution. Let

$$\Theta = \left|\left| \sum_{i=1}^{n} Y_i + \sum_{i=n+1}^{2n} Y_{i,n}^* \right|\right| \quad \text{and} \quad \Psi = \left|\left| \sum_{i=n+1}^{2n} \left(Y_i - Y_{i,n}^*\right) \right|\right|.$$

By applying the triangle inequality for the L^2 norm and for expected values, we get

$$\begin{aligned} E\left|\left|\sum_{i=1}^{2n} Y_i\right|\right|^{2+\delta} &= E\left|\left|\sum_{i=1}^{n} Y_i + \sum_{i=n+1}^{2n} Y_{i,n}^* + \sum_{i=n+1}^{2n}\left(Y_i - Y_{i,n}^*\right)\right|\right|^{2+\delta} \\ &\le E\left(\Theta + \Psi\right)^{2+\delta} \\ &\le \left((E\Theta^{2+\delta})^{1/(2+\delta)} + (E\Psi^{2+\delta})^{1/(2+\delta)}\right)^{2+\delta}. \end{aligned} \tag{2.35}$$

A two term Taylor expansion gives for all $a, b \ge 0$ and $r > 2$ that

$$(a+b)^r \le a^r + ra^{r-1}b + \frac{r(r-1)}{2}(a+b)^{r-2}b^2. \tag{2.36}$$

Since both of the expected values in the last line of the inequality in (2.35) are positive, we obtain by (2.36) that

$$\begin{aligned} E\left|\left|\sum_{i=1}^{2n} Y_i\right|\right|^{2+\delta} \le& E\Theta^{2+\delta} + (2+\delta)(E\Theta^{2+\delta})^{(1+\delta)/(2+\delta)}(E\Psi^{2+\delta})^{1/(2+\delta)} \\ &+ (2+\delta)(1+\delta)[(E\Theta^{2+\delta})^{1/(2+\delta)} \\ &+ (E\Psi^{2+\delta})^{1/(2+\delta)}]^{\delta}(E\Psi^{2+\delta})^{2/(2+\delta)}. \end{aligned} \tag{2.37}$$

We proceed by bounding the terms $(E\Psi^{2+\delta})^{1/(2+\delta)}$, and $E\Theta^{2+\delta}$ individually. Applications of both the triangle inequality for the L^2 norm and for expected values yield that

$$\begin{aligned} (E\Psi^{2+\delta})^{1/(2+\delta)} &= \left(E\left|\left|\sum_{i=n+1}^{2n}\left(Y_i - Y_{i,n}^*\right)\right|\right|^{2+\delta}\right)^{1/(2+\delta)} \\ &\le \left(E\left(\sum_{i=n+1}^{2n}||Y_i - Y_{i,n}^*||\right)^{2+\delta}\right)^{1/(2+\delta)} \\ &\le \sum_{i=n+1}^{2n}(E||Y_i - Y_{i,n}^*||^{2+\delta})^{1/(2+\delta)}. \end{aligned}$$

By Hölder's inequality, we have, with κ in (2.4),

$$(E||Y_i - Y^*_{i,n}||^{2+\delta})^{1/(2+\delta)} = (E[||Y_i - Y^*_{i,n}||^{(2+\delta)^2/\kappa}||Y_i - Y^*_{i,n}||^{(2+\delta)-(2+\delta)^2/\kappa}])^{1/(2+\delta)}$$
$$\leq (E||Y_i - Y^*_{i,n}||^{2+\delta})^{1/\kappa}(E||Y_i - Y^*_{i,n}||^{2+\delta})^{(\kappa-2-\delta)/(\kappa(2+\delta))}.$$

It follows from the definition of Y_i, $Y^*_{i,n}$ and the convexity of $x^{2+\delta}$ that

$$E||Y_i - Y^*_{i,n}||^{2+\delta} \leq 2^{1+\delta}(E||X_i - X^*_{i,n}||^{2+\delta} + E||X_{i,m} - Z_{i,n}||^{2+\delta}) \leq 2^{2+\delta}E||X_0 - X_{0,i-n}||^{2+\delta}$$

and

$$E||Y_i - Y^*_{i,n}||^{2+\delta} \leq 2^{1+\delta}(E||X_i - X_{i,m}||^{2+\delta} + E||X^*_{i,n} - Z_{i,n}||^{2+\delta}) \leq 2^{2+\delta}E||X_0 - X_{0,m}||^{2+\delta}.$$

Thus we get

$$(E\Psi^{2+\delta})^{1/(2+\delta)} \leq 2(E||X_0 - X_{0,m}||^{2+\delta})^{(\kappa-2-\delta)/(\kappa(2+\delta))}\sum_{\ell=1}^{\infty}(E||X_0 - X_{0,\ell}||^{2+\delta})^{1/\kappa} = J_m.$$

To bound $E\Theta^{2+\delta}$, since $\sum_{i=1}^{n} Y_i$ and $\sum_{i=n+1}^{2n} Y^*_{i,n}$ are independent and have the same distribution, we have by Lemma 2.3.2, Remark 2.3.1, and the inductive assumption that

$$\begin{aligned}
E\Theta^{2+\delta} &= E\left|\left|\sum_{i=1}^{n} Y_i + \sum_{i=n+1}^{2n} Y^*_{i,n}\right|\right|^{2+\delta} \\
&\leq 2E\left|\left|\sum_{i=1}^{n} Y_i\right|\right|^{2+\delta} + 2\left(E\left|\left|\sum_{i=1}^{n} Y_i\right|\right|^{2}\right)^{1+\delta/2} \\
&\leq 2n^{1+\delta/2}B + 2(nA)^{1+\delta/2}.
\end{aligned}$$

The substitution of these two bounds into (2.37) give that

$$\begin{aligned}
E\left|\left|\sum_{i=1}^{2n} Y_i\right|\right|^{2+\delta} &\leq 2n^{1+\delta/2}B + 2(nA)^{1+\delta/2} \\
&+ (2+\delta)[2n^{1+\delta/2}B + 2(nA)^{1+\delta/2}]^{(1+\delta)/(2+\delta)}J_m \\
&+ (2+\delta)(1+\delta)\left[2n^{1+\delta/2}B + 2(nA)^{1+\delta/2} + J_m\right]^{\delta} J_m^2.
\end{aligned} \tag{2.38}$$

Furthermore, by the definition of B, we may further bound each summand on the right-hand side of (2.38). We obtain for the first two terms that

$$\begin{aligned}
2n^{1+\delta/2}B + 2(nA)^{1+\delta/2} &\leq (2n)^{1+\delta/2}B\left[2^{-\delta/2} + \frac{A^{1+\delta/2}}{B}\right] \\
&\leq (2n)^{1+\delta/2}B\left[2^{-\delta/2} + 6c_\delta^{-1}\right].
\end{aligned}$$

A similar factoring procedure applied to the expression in the second line of (2.38) yields that

$$
\begin{aligned}
(2+\delta)&\left[2n^{1+\delta/2}B + 2(nA)^{1+\delta/2}\right]^{(1+\delta)/(2+\delta)} J_m \\
&\leq 6\left[(n^{1+\delta/2}B)^{(1+\delta)/(2+\delta)} + (nA)^{(1+\delta/2)[(1+\delta)/(2+\delta)]}\right] J_m \\
&\leq (2n)^{1+\delta/2}B\left[\frac{6J_m}{B^{1/(2+\delta)}} + \frac{6J_m A^{(1+\delta/2)[(1+\delta)/(2+\delta)]}}{B}\right] \\
&\leq (2n)^{1+\delta/2}B\left[12c_\delta^{-1}\right].
\end{aligned}
$$

Since $0 < \delta < 1$, the expression in the third line of (2.38) may be broken into three separate terms:

$$
\begin{aligned}
(2+\delta)(1+\delta)&\left[2n^{1+\delta/2}B + 2(nA)^{1+\delta/2} + J_m\right]^{\delta} J_m^2 \\
&\leq 6(2n^{1+\delta/2}B)^{\delta} J_m^2 + 6(2^{\delta}(nA)^{(1+\delta/2)\delta} J_m^2 + 6J_m^{2+\delta}.
\end{aligned}
$$

Furthermore, by again applying the definition of B, we have that

$$
\begin{aligned}
6(2n^{1+\delta/2}B)^{\delta} J_m^2 &= (2n)^{1+\delta/2}B\left[\frac{6(2n^{1+\delta/2}B)^{\delta} J_m^2}{(2n)^{1+\delta/2}B}\right] \\
&\leq (2n)^{1+\delta/2}B\left[\frac{6J_m^2}{B^{1-\delta}}\right] \\
&\leq (2n)^{1+\delta/2}B[6c_\delta^{-1}],
\end{aligned}
$$

$$
\begin{aligned}
6(2(nA)^{(1+\delta/2)})^{\delta} J_m^2 &= (2n)^{1+\delta/2}B\left[\frac{6(2(nA)^{(1+\delta/2)})^{\delta} J_m^2}{(2n)^{1+\delta/2}B}\right] \\
&\leq (2n)^{1+\delta/2}B\left[\frac{6A^{(1+\delta/2)\delta} J_m^2}{B}\right] \\
&\leq (2n)^{1+\delta/2}B[6c_\delta^{-1}],
\end{aligned}
$$

and

$$
6J_m^{2+\delta} = (2n)^{1+\delta/2}B\left[\frac{6J_m^{2+\delta}}{(2n)^{1+\delta/2}B}\right] \leq (2n)^{1+\delta/2}B\left[\frac{6J_m^{2+\delta}}{B}\right] \leq (2n)^{1+\delta/2}B[6c_\delta^{-1}].
$$

The application of these bounds to the right-hand side of (2.38) give that

$$
\begin{aligned}
E\left\|\sum_{i=1}^{2n} Y_i\right\|^{2+\delta} &\leq (2n)^{1+\delta/2}B\left[2^{-\delta/2} + 36c_\delta^{-1}\right] \\
&= (2n)^{1+\delta/2}B,
\end{aligned}
$$

which concludes the induction step and thus the proof. □

Theorem 2.3.2. *If (2.1)–(2.4) hold, then we have*

$$E\left(\max_{1\le k\le N}\left|\left|\sum_{i=1}^{k}(X_i - X_{i,m})\right|\right|\right)^{2+\delta} \le a_m N^{1+\delta/2} \tag{2.39}$$

with some sequence a_m satisfying $a_m \to 0$ as $m \to \infty$.

Proof. By examining the proofs, it is evident that Theorem 3.1 in Móricz et al.(1982) holds for L^2 valued random variables. Furthermore, by the stationarity of the sequence $\{X_i - X_{i,m}\}_{i=1}^{\infty}$ and Theorem 2.3.1, the conditions of Theorem 3.1 in Móricz are satisfied and therefore

$$E\left(\max_{1\le k\le N}\left|\left|\sum_{i=1}^{k}(X_i - X_{i,m})\right|\right|\right)^{2+\delta} \le c_\delta^* N^{1+\delta/2} B,$$

with some constant c_δ^*, depending only on δ and B is defined in (2.33). Observing that $B = B_m \to 0$ as $m \to \infty$, the result is proven. □

Theorem 2.3.1 provides inequality for the moments of the norm of partial sums of $X_i - X_{i,m}$ which are not Bernoulli shifts. However, checking the the proof of Theorem 2.3.1, we get the following result for Bernoulli shifts.

Theorem 2.3.3. *If (2.1), (2.3) are satisfied and* **X** *is a Bernoulli shift satisfying*

$$I(2+\delta) = \sum_{\ell=1}^{\infty}(E||X_0 - X_{0,\ell}||^{2+\delta})^{1/(2+\delta)} < \infty \quad \textit{with some } \ 0<\delta<1,$$

where $X_{0,\ell}$ is defined by (2.4), then for all $N \ge 1$

$$E\left|\left|\sum_{i=1}^{N} X_i\right|\right|^{2+\delta} \le N^{1+\delta/2} B_*,$$

where

$$B_* = E||X_0||^{2+\delta} + c_\delta^{2+\delta}[A_*^{1+\delta/2} + I^{2+\delta}(2+\delta) \\ + I(2+\delta)A_*^{(1+\delta)/2} + A_*^{(1+\delta/2)\delta} I^2(2+\delta)] + (c_\delta I^2(2))^{1/(1-\delta)},$$

$$A_* = \int EX_0^2(t)dt + 2\left(\int EX_0^2(t)dt\right)^{1/2} I(2)$$

and c_δ is defined in (2.34) and $I(2)$ in (2.30).

Remark 2.3.2. The inequality in Theorem 2.3.1 is an extension of Proposition 4 in Berkes et al.(2011) to random variables in Hilbert spaces; we have computed how B_* depends on the distribution of $\mathbf{X}$ explicitly.

Acknowledgement We are grateful to Professor Herold Dehling for pointing out several important references.

2.4 Bibliography

[1] Aue, A., Hörmann, S., Horváth, L., Hušková, M. and J. Steinebach, J.: Sequential testing for the stability of portfolio betas. *Econometric Theory*, **28**(2012), 804-837.

[2] Berkes, I. Hörmann, S. and Schauer, J.: Split invariance principles for stationary processes, *Annals of Probability*, **39**(2011), 2441-2473.

[3] Billingsley, P.:*Convergence of Probability Measures*, Wiley, New York, 1968.

[4] Bosq, D.: *Linear Processes in Function Spaces. Theory and Applications.* Lecture Notes in Statistics, 149. Springer-Verlag, New York, 2000.

[5] Bradley, R.C.: *Introduction to Strong Mixing Conditions I-III.* Kendrick Press, Heber City UT, 2007.

[6] Chow, Y. and Teicher, H.: *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag, Heidelberg, 1978.

[7] DasGupta, A.:*Asymptotic Theory of Statistics and Probability*, Springer, New York, 2008.

[8] Dedecker, J. Doukhan, P. Lang, G. León, R.J.R. Louhichi, S. and Prieur, C.:*Weak Dependence: with Examples and Applications.* Lecture Notes in Statistics, **190** Springer, New York, 2007. xiv+318 pp.

[9] Dedecker, J. and Merlevède, F.: The conditional central limit theorem in Hilbert spaces. *Stochastic Processes and Their Applications* **108** (2003), 229-262.

[10] Dehling, H.: Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **63** (1983), 393-432.

[11] Dehling, H. and Philip, W.: Almost sure invariance principles for weakly dependent vector-valued random variables.*Annals of Probability* **10** (1982), 689-701.

[12] Doob, J.: *Stochastic Processes* , Wiley, New York, 1953.

[13] Doukhan, P. and Louhichi, S.: A new weak dependence condition and applications to moment inequalities.*Stochastic Processes and Their Applications* **84** (1999), 313-342.

[14] Eberlein, E;: An invariance principle for lattices of dependent random variables. *Z. Wahrsch. Verw. Gebiete* **50**(1979), 119-133.

[15] Garsia, A.M.: Continuity properties of Gaussian processes with multidimensional time parameter. In: *Proceedings of the* 6^{th}*Berkeley Symposium Math. Stat. Probab.*, University of California, Berkeley, Vol. 2 (1970), pp. 369–374.

[16] Gordin, M.I.: The central limit theorem for stationary processes. (Russian) *Dokl. Akad. Nauk SSSR* **188** (1969), 739-741.

[17] Hardy, G.H., Littlewood, J.E. and Pólya, G.: *Inequalities.* (Second Edition) Cambridge University Press, 1959.

[18] Hörmann, S., Horváth, L. and Reeder, R.: *A functional version of the ARCH model. Econometric Theory*, 2012 In press.

[19] Hörmann, S. and Kokoszka, P.: Weakly dependent functional data. *Annals of Statistics* **38** (2010), 1845–1884.

[20] Horváth, L. and Kokoszka, P.: *Inference for Functional Data with Applications.* Springer, New York, 2012.

[21] Ibragimov, I. A.: Some limit theorems for stationary processes. *Theory of Probabability and Its Applications* **7** (1962), 349–382.

[22] Kuelbs, J. and Philipp, W.: Almost sure invariance principles for partial sums of mixing B-valued random variables. *Annals of Probabability* **8** (1980), 1003-1036.

[23] Merlevède, F.: Central limit theorem for linear processes with values in a Hilbert space. *Stochastic Processes and Their Applications* **65** (1996), 103-114.

[24] Móricz, F. Serfling, R. and Stout, W.: Moment and probability bounds with quasi-superadditive structure for the maximal partial sum, *Annals of Probability*, **10**(1982), 1032–1040.

[25] Philipp, W. and Stout, W.: Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoirs of the Amererican Mathematical Society* **161**(1975), iv+140 pp.

[26] Shorack, G. and Wellner, J.: *Empirical Processes With Applications To Statistics*, Wiley, New York, 1986.

[27] Wu, W.: Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** (2005), 14150-14154.

[28] Wu, W.: Strong invariance principles for dependent random variables. *Annals of Probability* **35** (2007), 2294-2320.

CHAPTER 3

TESTING STATIONARITY OF FUNCTIONAL TIME SERIES[2]

Economic and financial data often take the form of a collection of curves observed consecutively over time. Examples include, intraday price curves, yield and term structure curves, and intraday volatility curves. Such curves can be viewed as a time series of functions. A fundamental issue that must be addressed, before an attempt is made to statistically model such data, is whether these curves, perhaps suitably transformed, form a stationary functional time series. This chapter formalizes the assumption of stationarity in the context of functional time series and proposes several procedures to test the null hypothesis of stationarity. The tests are nontrivial extensions of the broadly used tests in the KPSS family. The properties of the tests under several alternatives, including change–point and $I(1)$, are studied, and new insights, present only in the functional setting, are uncovered. The theory is illustrated by a small simulation study and an application to intraday price curves.

3.1 Introduction

Over the last two decades, functional data analysis has become an important and steadily growing area of statistics. Very early on, major applications and theoretical developments pertained to functions observed consecutively over time, for example one function per day, or one function per year, with many of these data sets arising in econometric research. The main model employed for such series has been the functional autoregressive model of order one, which has received a great deal of attention; see Bosq (9), Antoniadis and Sapatinas (3), Antoniadis et al. (4), and Kargin and Onatski (30), among many others. More recent research has considered functional time series which have nonlinear dependence structure; see Hörmann and Kokoszka (21), Gabrys et al. (15), Horváth et al. (27), Hörmann et al. (23), as well as the review of Hörmann and Kokoszka (22) and Chapter 16 of Horváth and Kokoszka (25). As in traditional (scalar and vector) time series analysis, the underlying

[2] The content of this chapter is based on joint research with Piotr Kokoszka and Lajos Horváth.

assumption for inference in such models is stationarity. Stationarity is also required for functional dynamic regression models like those studied by Hays et al. (20) and Kokoszka et al. (33); for bootstrap and resampling methods for functional time series, see McMurry and Politis (37) and for the functional analysis of volatility, see Müller et al. (38).

Testing stationarity received due attention as soon as fundamental time series modeling principles have emerged. Early work includes Grenander and Rosenblatt (19), Granger and Hatanaka (18), and Priestley and Subba Rao (45). The methods considered by these authors rest on the spectral analysis which dominated the field of time series analysis at that time. While such approaches remain useful, see Dwivedi and Subba Rao (14), the spectral analysis of nonstationary functional time series has not been developed to a point where useable extensions could be readily derived. We note, however, the recent work of Panaretos and Tavakoli (39), Panaretos and Tavakoli (40) and Hörmann et al. (24), who advance the spectral analysis of stationary functional time series.

We follow a time domain approach introduced in the seminal paper of Kwiatkowski et al. (34) which is now firmly established in econometric theory and practice, and has been extended in many directions. The work of Kwiatkowski et al. (34) was motivated by the fact that unit root tests developed by Dickey and Fuller (11), Dickey and Fuller (12), and Said and Dickey (47) indicated that most aggregate economic series had a unit root. In these tests, the null hypothesis is that the series has a unit root. Since such tests have low power in samples of sizes occurring in many applications, Kwiatkowski et al. (34) proposed that stationarity should be considered as the null hypothesis (they used a broader definition which allowed for deterministic trends), and the unit root should be the alternative. Rejection of the null of stationarity could then be viewed as a convincing evidence in favor of a unit root. It was soon realized that the KPSS test of Kwiatkowski et al. (34) has a much broader utility. For example, Lee and Schmidt (35) and Giraitis et al. (17) used it to detect long memory, with short memory as the null hypothesis. At present, both the augmented Dickey–Fuller test and the KPSS test, as well as its robust version of de Jong et al. (10), are typically applied to the same series to get a fuller picture. They are available in many packages, including `R` and `Matlab` implementations. The work of Lo (36) is also very relevant to our approach. His contribution is crucial because he showed that to obtain parameter free limit null distributions, statistics similar to the KPSS statistic must be normalized by the long–run variance rather than by the sample variance, which leads to these distributions only if the observations are independent.

This chapter seeks to develop a general methodology for testing the assumption that a functional time series to be modeled is indeed stationary and weakly dependent. Such a test should be applied before fitting one of the known stationary models (all of them are weakly dependent). In many cases, it will be applied to functions transformed to remove seasonality or obvious trends, or to model residuals. At present, only CUSUM change point tests are available for functional time series; see Berkes et al. (7), Horváth et al. (26), and Zhang et al. (53). These tests have high power to detect abrupt changes in the stochastic structure of a functional time series, either the mean or the covariance structure. Our objective is to develop more general tests of stationarity which also have high power against integrated and other alternatives.

It is difficult to explain the main contribution of this chapter without introducing the required notation, but we wish to highlight in this paragraph the main difficulties which are encountered in the transition from the scalar or vector to the functional case. A stationary functional time series can be represented as

$$X_n(t) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_{jn}v_j(t),$$

where n is the time index that counts the functions (referring e.g. to a day), and t is the (theoretically continuous) argument of each function. The mean function μ and the functional principal components v_j are unknown deterministic functions which depend on the stochastic structure of the series $\{X_n\}$, and which are estimated by random functions $\hat{\mu}$ and $\hat{v}_j$. If $\{X_n\}$ is not stationary, one can still compute the estimators $\hat{\mu}$ and $\hat{v}_j$, but they will not converge to μ or v_j because these population quantities will not exist then. Thus the use of a data–driven basis system v_j represents an aspect which is not encountered in the theory of scalar or vector valued tests. Therefore, after defining meaningful extensions to the functional setting, we must develop a careful analysis of the behavior of the tests under alternatives.

The chapter is organized as follows. Section 3.2 formalizes the null hypothesis of stationarity and weak dependence of functional time series, introduces the tests, and explores their asymptotic properties under the null hypothesis. In Section 3.3, we turn to the behavior of the tests under several alternatives. Section 3.4 explains the details of the implementation, and contains the results of a simulation study, while Section 3.5 illustrates the properties of the tests by an application to intraday price curves. Appendices 3.6 and 3.7 contain, respectively, the proofs of the results stated in Sections 3.2 and 3.3.

3.2 Assumptions and test statistics

Linear functional time series, in particular functional AR(1) processes, have the form $X_n = \sum_j \Psi_j(\varepsilon_{n-j})$, where the ε_i are iid error functions, and the Ψ_j are bounded linear operators acting on the space of square integrable functions. In this chapter, we assume merely that $X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots)$, for some, possibly nonlinear, function f. The operators Ψ_j or the function f arise as solutions to structural equations, very much like in the univariate econometric modeling; see e.g. Teräsvirta et al. (50). For the functional autoregressive process, the norms of the operators Ψ_j decay exponentially fast. For the more general nonlinear moving averages, the rate at which the dependence of X_n on past errors ε_{n-j} decays with j can be quantified by a condition known as L^p–m–approximability stated in assumptions (3.1)–(3.4) below. In both cases, these functional models can be said to be in a class which is customarily referred to as weakly dependent or short memory time series. It is convenient to state the conditions for the error process, which we denote by $\boldsymbol{\beta} = \{\eta_j\}_{-\infty}^{\infty}$, and which will be used to formulate the null and alternative hypotheses.

Throughout the chapter, L^2 denotes the Hilbert space of square integrable functions on the unit interval $[0,1]$ with the usual inner product $\langle\cdot,\cdot\rangle$ and the norm $||\cdot||$ it generates, $\int$ means $\int_0^1$.

$$\begin{aligned}&\boldsymbol{\beta} \text{ forms a sequence of Bernoulli shifts, i.e. } \eta_j = g(\varepsilon_j, \varepsilon_{j-1}, \ldots)\\ &\text{for some measurable function } g: S^\infty \mapsto L^2 \text{ and iid functions } \varepsilon_j,\\ &-\infty < j < \infty, \text{with values in a measurable space } S,\end{aligned} \tag{3.1}$$

$$\varepsilon_j(t) = \varepsilon_j(t,\omega) \text{ is jointly measurable in } (t,\omega), \quad -\infty < j < \infty, \tag{3.2}$$

$$E\eta_0(t) = 0 \text{ for all } t, \text{ and } E||\eta_0||^{2+\delta} < \infty, \quad \text{for some} \quad 0 < \delta < 1, \tag{3.3}$$

and

$$\begin{aligned}&\text{the sequence } \{\eta_n\}_{n=-\infty}^{\infty} \text{ can be approximated by } \ell\text{–dependent}\\ &\text{sequences } \{\eta_{n,\ell}\}_{n=-\infty}^{\infty} \text{ in the sense that}\\ &\qquad\sum_{\ell=1}^{\infty} (E||\eta_n - \eta_{n,\ell}||^{2+\delta})^{1/\kappa} < \infty \text{ for some } \kappa > 2+\delta,\\ &\text{where } \eta_{n,\ell} \text{ is defined by } \eta_{n,\ell} = g(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-\ell+1}, \boldsymbol{\epsilon}^*_{n,\ell}),\\ &\boldsymbol{\epsilon}*_{n,\ell} = (\varepsilon^*_{n,\ell,n-\ell}, \varepsilon^*_{n,\ell,n-\ell-1}, \ldots), \text{where the } \varepsilon^*_{n,\ell,k}\text{'s are independent copies of } \varepsilon_0,\end{aligned} \tag{3.4}$$

independent of $\{\varepsilon_i, -\infty < i < \infty\}$.

Assumptions similar to those stated above have been used extensively in recent theoretical work, as all stationary time series models in practical use can be represented as Bernoulli shifts; see Wu (52), Shao and Wu (48), Aue et al. (5), Hörmann and Kokoszka (21), among many other contributions. They have been used in econometric research even earlier, and the work of Pötscher and Prucha (44) contributed to their popularity. Bernoulli shifts are stationary by construction; weak dependence is quantified by the summability condition in (3.4) which intuitively states that the function g decays so fast that the impact of shocks far back in the past is so small that they can be replaced by their independent copies, with only a small change in the distribution of the process.

We wish to test

$$H_0: \quad X_i(t) = \mu(t) + \eta_i(t), \quad 1 \le i \le N, \quad \text{where} \quad \mu \in L^2.$$

The mean function μ is unknown. The null hypothesis is that the functional time series is stationary and weakly dependent, with the structure of dependence quantified by conditions (3.1)–(3.4).

The most general alternative is that H_0 does not hold, but some profound insights into the behavior of the tests can be obtained by considering some specific alternatives. We focus on the following.

Change point alternative:

$$H_{A,1}: \quad X_i(t) = \mu(t) + \delta(t)I\{i > k^*\} + \eta_i(t), \; 1 \le i \le N, \text{with some integer } \; 1 \le k^* < N.$$

The mean function $\mu(t)$, the size of the change $\delta(t)$, and the time of the change, k^*, are all unknown parameters. We assume that the change occurs away from the end points, i.e.

$$k^* = \lfloor N\tau \rfloor \quad \text{with some} \quad 0 < \tau < 1. \tag{3.5}$$

Integrated alternative:

$$H_{A,2}: \quad X_i(t) = \mu(t) + \sum_{\ell=1}^{i} \eta_\ell(t), \; 1 \le i \le N.$$

Deterministic trend alternative:

$$H_{A,3}: \quad X_i(t) = \mu(t) + g(i/N)\delta(t) + \eta_i(t), \; 1 \le i \le N \tag{3.6}$$

where

$$g(t) \quad \text{is a piecewise Lipschitz continuous function on } [0,1]. \tag{3.7}$$

The trend alternative includes various change point alternatives, including $H_{A,1}$, but also those in which change can be gradual. It also includes the polynomial trend alternative, if $g(u) = u^{\alpha}$.

We emphasize that both under the null hypothesis and all alternatives, the mean function $\mu(t)$ is unknown.

The tests we propose can be shown to be consistent against any other sufficiently large departures from stationarity and weak dependence. In particular, functional long memory alternatives could be considered as well, as studied in the scalar case by Giraitis et al. (17). Since long memory functional processes have not been considered in any applications yet, we do not pursue this direction at this point.

In the remainder of this section, we consider two classes of tests, those based on the curves themselves, and those based on the finite dimensional projections of the curves on the functional principal components. As will become clear, the tests of the two types are related.

3.2.1 Fully functional tests

Our approach is based on two tests statistics. The first is

$$T_N = \iint Z_N^2(x,t)dtdx,$$

where

$$Z_N(x,t) = S_N(x,t) - xS_N(1,t), \quad 0 \le x, t \le 1,$$

with

$$S_N(x,t) = N^{-1/2} \sum_{i=1}^{\lfloor Nx \rfloor} X_i(t), \quad 0 \le x, t \le 1.$$

The second test statistic is

$$M_N = T_N - \int \left(\int Z_N(x,t)dx \right)^2 dt = \iint \left(Z_N(x,t) - \int Z_N(y,t)dy \right)^2 dxdt.$$

If $X_i(t) = X_i$, i.e. if the data are scalars (or constant functions on $[0,1]$), the statistic T_N is the numerator of the KPSS statistic of Kwiatkowski et al. (34), and M_N is the numerator of the V/S statistic of Giraitis et al. (17), who introduced centering to reduce the variability of the KPSS statistic and to increase power against "changes in variance" which are a characteristic of long memory in volatility. As pointed out by Lo (36), to obtain parameter free limits under the null, statistics of this type must be divided by the long–run variance.

We now proceed with the suitable definitions in the functional case.

The null limit distributions of T_N and M_N depend on the eigenvalues of the long–run covariance function of the errors:

$$C(t,s) = E\eta_0(t)\eta_0(s) + \sum_{\ell=1}^{\infty} E\eta_0(t)\eta_\ell(s) + \sum_{\ell=1}^{\infty} E\eta_0(s)\eta_\ell(t). \tag{3.8}$$

It is proven in Horváth et al. (27) that the series in (3.8) is convergent in L^2. The function $C(t,s)$ is positive definite, and therefore, there exist $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and orthonormal functions $\varphi_i(t), 0 \leq t \leq 1$, satisfying

$$\lambda_i \varphi_i(t) = \int C(t,s)\varphi_i(s)ds, \quad 1 \leq i < \infty. \tag{3.9}$$

The following theorem specifies limit distributions of T_N and M_N under the stationarity null hypothesis. Throughout the chapter, $B_1, B_2, \ldots$ are independent Brownian bridges.

Theorem 3.2.1. *If assumptions* (3.1)–(3.4) *and* H_0 *hold, then*

$$T_N \quad \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \lambda_i \int B_i^2(x)dx \tag{3.10}$$

and

$$M_N \quad \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \lambda_i \int \left(B_i(x) - \int B_i(y)dy \right)^2 dx. \tag{3.11}$$

According to Theorem 3.6.1, under assumptions (3.1)–(3.4), the sum $\sum_{i=1}^{\infty} \lambda_i$ is finite, and therefore, the variables T_0 and M_0 are finite with probability one.

Theorem 3.2.1 shows, in particular, that for functional time series, a simple normalization with a long–run variance is not possible, and approaches involving the estimation of all large eigenvalues must be employed. The eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ can be easily estimated under the null hypothesis because then

$$C(t,s) = \operatorname{cov}(X_0(t), X_0(s)) + \sum_{i=1}^{\infty} [\operatorname{cov}(X_0(t), X_i(s)) + \operatorname{cov}(X_0(s), X_i(t))],$$

so we can use the kernel estimator $\hat{C}_N$ of Horváth et al. (27) defined as

$$\hat{C}_N(t,s) = \hat{\gamma}_0(t,s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)(\hat{\gamma}_i(t,s) + \hat{\gamma}_i(s,t)), \tag{3.12}$$

where

$$\hat{\gamma}_i(t,s) = \frac{1}{N} \sum_{j=i+1}^{N} \left(X_j(t) - \bar{X}_N(t)\right)\left(X_{j-i}(s) - \bar{X}_N(s)\right)$$

with

$$\bar{X}_N(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t).$$

The kernel K in the definition of $\hat{C}_N$ satisfies the following conditions:

$$K(0) = 1, \tag{3.13}$$

$$K(u) = 0 \text{ if } u > c \text{ with some } c > 0, \tag{3.14}$$

and

$$K \text{ is continuous on } [0, c], \text{ where } c \text{ is given in (3.14).} \tag{3.15}$$

The window (or smoothing bandwidth) h must satisfy only

$$h = h(N) \to \infty \quad \text{and} \quad \frac{h(N)}{N} \to 0, \quad \text{as} \quad N \to \infty. \tag{3.16}$$

Now the estimators for the eigenvalues and eigenfunctions are defined by

$$\hat{\lambda}_i \hat{\varphi}_i(t) = \int \hat{C}_N(t, s)\hat{\varphi}_i(s)ds, \;\; 1 \le i \le N,$$

where $\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \ldots$ are the empirical eigenvalues and $\hat{\varphi}_1, \hat{\varphi}_2, \ldots$ are the corresponding orthonormal eigenfunctions. We can thus approximate the limits in Theorem 3.2.1 with

$$\sum_{i=1}^{d} \hat{\lambda}_i \int B_i^2(x)dx \quad \text{and} \quad \sum_{i=1}^{d} \hat{\lambda}_i \int \left(B_i(x) - \int B_i(y)dy\right)^2 dx,$$

where d is suitably large. The details are presented in Section 3.4. We note that the $\hat{\lambda}_i$ and the $\hat{\varphi}_i$ are consistent estimators only under H_0. Their behavior under the alternatives is complex. It is studied in Section 3.3.

3.2.2 Tests based on projections

Theorem 3.2.1 leads to asymptotic distributions depending on the eigenvalues λ_i, which can collectively be viewed as an analog of the long–run variance. In this section, we will see that by projecting on the eigenfunctions $\hat{\varphi}_i$, it is possible to construct statistics whose limit null distributions are parameter free. This procedure is a functional analog of dividing by an estimator of a long–run variance.

To have uniquely defined (up to the sign) eigenfunctions, we assume

$$\lambda_1 > \lambda_2 > \ldots \lambda_d > \lambda_{d+1} > 0. \tag{3.17}$$

Define

$$T_N^0(d) = \sum_{i=1}^{d} \frac{1}{\hat{\lambda}_i} \int \langle Z_N(x,\cdot), \hat{\varphi}_i \rangle^2 dx,$$

$$T_N^*(d) = \sum_{i=1}^{d} \int \langle Z_N(x,\cdot), \hat{\varphi}_i \rangle^2 dx,$$

$$M_N^0(d) = \sum_{i=1}^{d} \frac{1}{\hat{\lambda}_i} \int \left(\langle Z_N(x,\cdot), \hat{\varphi}_i \rangle - \int \langle Z_N(u,\cdot), \hat{\varphi}_i \rangle du \right)^2 dx$$

and

$$M_N^*(d) = \sum_{i=1}^{d} \int \left(\langle Z_N(x,\cdot), \hat{\varphi}_i \rangle - \int \langle Z_N(u,\cdot), \hat{\varphi}_i \rangle du \right)^2 dx.$$

Theorem 3.2.2. *If assumptions* (3.1)–(3.4), (3.13)–(3.16), (3.17) *and H_0 hold, then*

$$T_N^0(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^{d} \int B_i^2(x) dx, \tag{3.18}$$

$$T_N^*(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^{d} \lambda_i \int B_i^2(x) dx, \tag{3.19}$$

$$M_N^0(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^{d} \int \left(B_i(x) - \int B_i(u) du \right)^2 dx \tag{3.20}$$

and

$$M_N^*(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^{d} \lambda_i \int \left(B_i(x) - \int B_i(u) du \right)^2 dx. \tag{3.21}$$

It is clear that T_N^* and M_N^* are just d–dimensional projections of T_N and M_N. The distribution of the limit in (3.18) can be found in Kiefer (31). Critical values based on Monte Carlo simulations are given in Table 6.1 of Horváth and Kokoszka (25). The distributions of the limits both in (3.18) and (3.20) can also be expressed in terms of sums of squared normals; see Shorack and Wellner (49) and Section 3.4. It is also easy to derive normal approximations. By the central limit theorem, we have, as $d \to \infty$,

$$\left(\frac{45}{d} \right)^{-1/2} \left[\sum_{i=1}^{d} \int B_i^2(x) dx - \frac{d}{6} \right] \quad \xrightarrow{\mathcal{D}} \quad N(0,1),$$

where $N(0,1)$ stands for a standard normal random variable. Aue et al. (5) demonstrated that the limit in (3.18) can be approximated well with normal random variables even for moderate d. The limit in (3.20) can be approximated in a similar manner, namely, as $d \to \infty$,

$$\left(\frac{360}{d}\right)^{-1/2}\left[\sum_{i=1}^{d}\left\{\int B_i^2(x)dx - \left(\int B_i(x)dx\right)^2\right\} - \frac{d}{12}\right] \quad \xrightarrow{\mathcal{D}} \quad N(0,1).$$

3.3 Asymptotic behavior under alternatives

The asymptotic behavior of the KPSS and related tests under alternatives is not completely understood, even for scalar data. This may be due to the fact that an asymptotic analysis of power is generally much more difficult than the theory under a null hypothesis. Giraitis et al. (17) studied the behavior of the KPSS test, the R/S test of Lo (36), and their V/S test under the alternative of long memory. Pelagatti and Sen (41) established the consistency of their nonparametric version of the KPSS test under the integrated alternative. In this section, we present an asymptotic analysis, under alternatives, of the tests introduced in Section 3.2. In the functional setting, there is a fundamentally new aspect: convergence of a scalar estimator of the long–run variance must be replaced by the convergence of the eigenvalues and the eigenfunctions of the long–run covariance function. We derive precise rates of convergence and limits for this function, and use them to study the asymptotic power of the tests introduced in Section 3.2. In Section 3.4, we will see how these asymptotic insights manifest themselves in finite samples.

We expect that the tests introduced in Section 3.2 are also consistent against suitably defined long memory alternatives. While scalar long memory models have received a lot of attention in recent decades, long memory functional models have not been considered in econometric literature yet. To keep this contribution within reasonable limits, we do not pursue this direction here.

3.3.1 Change in the mean alternative

To state consistency results, we assume that the jump function is in L^2, i.e.

$$\int \delta^2(t)dt < \infty. \tag{3.22}$$

We introduce the function

$$\delta_\tau(x,t) = \delta(t)[(x-\tau)I\{x \geq \tau\} - x(1-\tau)] \tag{3.23}$$

and the Gaussian process $\Gamma^0(x,t)$ with $E\Gamma^0(x,t) = 0$ and

$$E\Gamma^0(x,t)\Gamma^0(y,s) = (\min(x,y) - xy)C(t,s).$$

The existence of the process $\Gamma^0(x,t)$ will be established in Appendix 3.6.

Theorem 3.3.1. *If assumptions* (3.1)–(3.4), (3.5), (3.22), *and* $H_{A,1}$ *hold, then*

$$N^{-1/2}\left\{T_N - \frac{N}{3}\tau^2(1-\tau)^2||\delta||^2\right\} \xrightarrow{\mathcal{D}} 2\iint \Gamma^0(x,t)\delta_\tau(x,t)dtdx \tag{3.24}$$

and

$$\begin{aligned} &N^{-1/2}\left\{M_N - \frac{N}{12}\tau^2(1-\tau)^2||\delta||^2\right\} \\ &\xrightarrow{\mathcal{D}} 2\iint \left(\Gamma^0(x,t) - \int \Gamma^0(y,t)dy\right)\left(\delta_\tau(x,t) - \int \delta_\tau(y,t)dy\right)dtdx. \end{aligned} \tag{3.25}$$

It is easy to see that the limits in Theorem 3.3.1 are zero mean normal random variables. Their variances, computed in Appendix 3.7, are positive if $C(t,s)$ is strictly positive definite. In that case, T_N and M_N increase like N. However, as we prove in Lemma 3.7.2, $\hat{C}_N(t,s)$ does not converge to $C(t,s)$ under $H_{A,1}$, so it is not clear what the asymptotic behavior of the critical values under $H_{A,1}$ is. To show that the asymptotic power is 1, a more delicate argument is needed, which we now outline.

Applying Lemma 3.7.2 with the result of Dunford and Schwartz (13), p. 1091, we conclude that

$$\frac{\hat{\lambda}_1}{h} \xrightarrow{P} \gamma_{A,1} = 2\tau(1-\tau)||\delta||^2\int_0^c K(u)du, \tag{3.26}$$

and

$$\left|\left|\hat{\varphi}_1(t) - \hat{c}_1\frac{\delta(t)}{||\delta||}\right|\right| = o_P(1). \tag{3.27}$$

According to (3.26), when we compute $\bar{c} = \bar{c}(h,N)$, the critical value from simulated copies of $\sum_{i=1}^d \hat{\lambda}_i \int B_i^2(t)dt$, then $\bar{c}$ increases at most linearly with h. Therefore, using (3.16) with Theorem 3.3.1, we conclude that

$$\lim_{N\to\infty} P\{T_N \geq \bar{c}\} = 1 \quad \text{under} \quad H_{A,1}. \tag{3.28}$$

This shows that the test based on T_N is consistent. The same argument applies to M_N.

We now turn to the tests based on projections, with the test statistics defined in Section 3.2.2. As we have seen, under $H_{A,1}$, the largest empirical eigenvalue $\hat{\lambda}_1$ increases to ∞, as $N \to \infty$, and the corresponding empirical eigenfunction $\hat{\varphi}_1$ is asymptotically in the direction of the change. This means that both T_N^* and M_N^* are dominated by the first term under $H_{A,1}$. The precise asymptotic behavior of all statistics introduced in Section 3.2.2 is described in the following theorem.

Theorem 3.3.2. *If assumptions* (3.1)–(3.4), (3.13)–(3.16), (3.22), *and* $H_{A,1}$ *hold, then*

$$N^{-1/2}\left\{T_N^*(1)-\frac{N}{3}\tau^2(1-\tau)^2\langle\delta,\hat{\varphi}_1\rangle^2\right\} \quad \xrightarrow{\mathcal{D}} \quad 2\iint \Gamma^0(x,t)\delta_\tau(x,t)dxdt, \tag{3.29}$$

$$\begin{aligned} &N^{-1/2}\left\{M_N^*(1)-\frac{N}{12}\tau^2(1-\tau)^2\langle\delta,\hat{\varphi}_1\rangle^2\right\} \\ &\quad \xrightarrow{\mathcal{D}} \quad 2\iint\left(\Gamma^0(x,t)-\int\Gamma^0(y,t)dy\right)\left(\delta_\tau(x,t)-\int\delta_\tau(y,t)dy\right)dtdx, \end{aligned} \tag{3.30}$$

$$\begin{aligned} &\frac{h}{N^{1/2}}2\tau(1-\tau)||\delta||^2\int_0^c K(u)du\left\{T_N^0(1)-\frac{N}{3\hat{\lambda}_1}\tau^2(1-\tau)^2\langle\delta,\hat{\varphi}_1\rangle^2\right\} \\ &\qquad \xrightarrow{\mathcal{D}} \quad 2\iint\Gamma^0(x,t)\delta_\tau(x,t)dxdt, \end{aligned} \tag{3.31}$$

and

$$\begin{aligned} &\frac{h}{N^{1/2}}2\tau(1-\tau)||\delta||^2\int_0^c K(u)du\left\{M_N^0(1)-\frac{N}{12\hat{\lambda}_1}\tau^2(1-\tau)^2\langle\delta,\hat{\varphi}_1\rangle^2\right\} \\ &\xrightarrow{\mathcal{D}} \quad 2\iint\left(\Gamma^0(x,t)-\int\Gamma^0(y,t)dy\right)\left(\delta_\tau(x,t)-\int\delta_\tau(y,t)dy\right)dtdx. \end{aligned} \tag{3.32}$$

If in addition we assume that $h/N^{1/2}\to 0$ *as* $N\to\infty$*, then*

$$T_N^*(d)=\frac{N}{3}\tau^2(1-\tau)^2||\delta||^2(1+o_P(1)), \tag{3.33}$$

$$M_N^*(d)=\frac{N}{12}\tau^2(1-\tau)^2||\delta||^2(1+o_P(1)), \tag{3.34}$$

$$T_N^0(d)=\frac{N}{h}\frac{\tau(1-\tau)}{6\int_0^c K(u)du}(1+o_P(1)), \tag{3.35}$$

and

$$M_N^0(d)=\frac{N}{h}\frac{\tau(1-\tau)}{24\int_0^c K(u)du}(1+o_P(1)). \tag{3.36}$$

Observe that according to Theorems 3.3.1 and 3.3.2, the statistics T_N and $T_N^*(1)$ (M_N and $M_N^*(1)$, respectively) exhibit the same asymptotic behavior under the change point alternative. This is due to the fact that the projection in the direction of $\hat{\varphi}_1$ picks up all information on the change available in the data, as, by (3.27), $\hat{\varphi}_1$ is asymptotically aligned with the direction of the change.

Remark 3.1. In the local alternative model

$$X_i(t) = \mu(t) + \delta_N^*(t)I\{i > k^*\} + \eta_i(t),\ 1 \le i \le N, \text{with some integer } \ 1 \le k^* < N,$$

where $||\delta_N^*|| \to 0$ as $N \to \infty$. We discuss briefly how the statistic T_N behaves under this model. If $N^{1/2}||\delta_N^*|| \to 0$, then T_N converges in distribution to $\iint (\Gamma^0(x,t))^2 dtdx$ as is the case under H_0. On the other hand, if $N^{1/2}||\delta_N^*|| \to \infty$, then $T_N \xrightarrow{P} \infty$ and therefore, consistency is retained. Moreover, under the additional assumption $N \iint C(t,s)\delta_N^*(t)\delta_N^*(s)dtds \to \infty$ we show that

$$\frac{1}{A_N}\left\{T_N - \frac{1}{N}||\beta_N||^2\right\} \xrightarrow{\mathcal{D}} N(0,1), \tag{3.37}$$

where

$$A_N^2 = 4N \iint C(t,s)\delta_N^*(t)\delta_N^*(s)dtds \iint (\min(x,y) - xy)\bar{\delta}_\tau(x)\bar{\delta}_\tau(y)dxdy.$$

In the critical case when $N^{1/2}||\delta_N^*|| \to \delta^*$ in L^2, where δ^* is some non zero function, then we have

$$T_N \xrightarrow{\mathcal{D}} \zeta + \sum_{\ell=1}^{\infty} \left\{\lambda_\ell ||B_\ell||^2 + 2\lambda_\ell^{1/2}\langle B_\ell, \bar{\delta}_\tau\rangle\langle\varphi_\ell, \delta^*\rangle\right\}, \tag{3.38}$$

where $\zeta = ||\bar{\delta}_\tau||^2||\delta^*||^2$, $B_1, B_2, \ldots$ are independent Brownian bridges, the λ_i's and φ's are defined in (3.9), and

$$\bar{\delta}_\tau(x) = (x-\tau)I\{x \ge \tau\} - x(1-\tau). \tag{3.39}$$

The asymptotic behavior of M_N can be studied analogously in the local alternative change point model. The derivation of the asymptotic properties of $T_N^0(d), T_N^*(d), M_N^0(d),$ and $M_N^*(d)$ is much more involved since it requires the study of $\hat{C}_N(t,s)$ under this model. We will not pursue this line of inquiry in the present chapter.

3.3.2 The integrated alternative

Let

$$\Delta(x,t) = \int_0^x \Gamma(u,t)du - x\int \Gamma(u,t)du, \tag{3.40}$$

where $\Gamma(x,t)$ is a Gaussian process with $E\Gamma(x,t) = 0$ and $E\Gamma(x,t)\Gamma(y,s) = \min(x,y)C(t,s)$. The existence of $\Gamma(x,t)$ is established in Theorem 3.6.1.

For the fully functional tests of Section 3.2.1, we have the following result.

Theorem 3.3.3. *If assumptions* (3.1)–(3.4) *and* $H_{A,2}$ *hold, then*

$$\frac{1}{N^2}T_N \quad \xrightarrow{\mathcal{D}} \quad \iint \Delta^2(x,t)dtdx \tag{3.41}$$

and

$$\frac{1}{N^2}M_N \quad \xrightarrow{\mathcal{D}} \quad \iint \left(\Delta(x,t) - \int \Delta(u,t)du\right)^2 dtdx. \tag{3.42}$$

To find the limit distributions of the statistics based on projections, we need the following theorem.

Theorem 3.3.4. *If assumptions* (3.1)–(3.4), (3.13)–(3.16), *and* $H_{A,2}$ *hold, then*

$$\left\{\frac{1}{N}Z_N(x,t), \frac{1}{Nh}\hat{C}_N(t,s), 0 \le x,t,s \le 1\right\} \longrightarrow \left\{\Delta(x,t), Q(t,s), 0 \le x,t,s \le 1\right\}$$

in $\mathcal{D}([0,1] \times L^2)$, *where*

$$Q(t,s) = 2\left(\int_0^c K(w)dw\right)\int R(z,t)R(z,s)dz,$$

with

$$R(z,t) = \int_0^z \Gamma(u,t)du - \int \left\{\int_0^v \Gamma(u,t)du\right\} dv.$$

We show in Lemma 3.7.5 that $Q(t,s)$ is non–negative definite with probability one, so there are random variables $\lambda_1^* \ge \lambda_2^* \ge \dots$ and random functions $\varphi_1^*(t), \varphi_2^*(t), \dots$ satisfying

$$\lambda_i^* \varphi_i^*(t) = \int Q(t,s)\varphi_i^*(s)ds, \quad 1 \le i < \infty. \tag{3.43}$$

Combining Theorem 3.3.4 with Dunford and Schwartz (13), we get that

$$\begin{aligned}&\left(\hat{\lambda}_1/(Nh), \hat{\lambda}_2/(Nh), \dots, \dots, \hat{\lambda}_d/(Nh), \hat{\varphi}_1(t), \hat{\varphi}_2(t), \dots, \hat{\varphi}_d(t)\right)\\ &\quad \xrightarrow{\mathcal{D}} \left(\lambda_1^*, \lambda_2^*, \dots, \lambda_d^*, \varphi_1^*(t), \varphi_2^*(t), \dots, \varphi_d^*(t)\right).\end{aligned}$$

Thus the behavior of $T_N^0(d), T_N^*(d), M_N^0(d)$ and $M_N^*(d)$ is an immediate consequence of Theorem 3.3.4. An argument similar to that developed in Section 3.3.1 shows that the tests are consistent.

Theorem 3.3.5. *If assumptions* (3.1)–(3.4), (3.13)–(3.16), *and* $H_{A,2}$ *hold, then*

$$\frac{h}{N}T_N^0(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^d \frac{1}{\lambda_i^*}\int \langle \Delta(x,\cdot), \varphi_i^*(\cdot)\rangle^2 dx, \tag{3.44}$$

$$\frac{1}{N^2}T_N^*(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^d \int \langle \Delta(x,\cdot), \varphi_i^*(\cdot)\rangle^2 dx, \tag{3.45}$$

$$\frac{h}{N}M_N^0(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^d \frac{1}{\lambda_i^*}\int \left(\langle \Delta(x,\cdot), \varphi_i^*(\cdot)\rangle - \int \langle \Delta(u,\cdot), \varphi_i^*(\cdot)\rangle du\right)^2 dx, \tag{3.46}$$

and

$$\frac{1}{N^2}M_N^*(d) \quad \xrightarrow{\mathcal{D}} \quad \sum_{i=1}^{d}\int\left(\langle\Delta(x,\cdot),\varphi_i^*(\cdot)\rangle - \int\langle\Delta(u,\cdot),\varphi_i^*(\cdot)\rangle du\right)^2 dx. \tag{3.47}$$

3.3.3 Deterministic trend alternative

Let

$$\bar{g}(x) = \int_0^x g(u)du - x\int g(u)du, \qquad 0 \le x \le 1.$$

Theorem 3.3.6. *If assumptions* (3.1)–(3.4), (3.6), (3.22) *and* $H_{A,3}$ *hold, then*

$$N^{-1/2}\left\{T_N - N||\delta||^2\int \bar{g}^2(x)dx\right\} \quad \xrightarrow{\mathcal{D}} \quad 2\iint \Gamma^0(x,t)\delta(t)\bar{g}(x)dtdx \tag{3.48}$$

and

$$N^{-1/2}\left\{M_N - N||\delta||^2\int\left(\bar{g}(x) - \int\bar{g}(y)dy\right)^2 dx\right\} \tag{3.49}$$

$$\xrightarrow{\mathcal{D}} \quad 2\iint\left(\Gamma^0(x,t) - \int\Gamma^0(y,t)dy\right)\left(\bar{g}(x) - \int\bar{g}(y)dy\right)\delta(t)dtdx.$$

The limits in (3.48) and (3.49) are normal random variables with zero mean and variances which can be expressed in terms of the long–run covariance kernel $C(\cdot,\cdot)$ and the functions δ and $\bar{g}$. We do not display these complex formulas to conserve space. They extend the formulas for the variances of the limits in Theorem 3.3.1 which are given in Appendix 3.7. The consistency of the procedures based on projections can be established by extending the arguments used to prove Theorem 3.3.2, however with more abstract notation. Again, to keep this work within reasonable limits of space, we do not present the details.

3.4 Implementation and finite sample performance

In this section we discuss the implementation of the testing procedure developed in the sections above. A simulation study is then presented in order to investigate the finite sample properties of the test.

3.4.1 Details of the implementation

To implement the tests introduced in Section 3.2, several issues must be considered. The choice of the kernel $K(\cdot)$ and the smoothing bandwidth h are the most obvious. Beyond that, to implement Monte Carlo tests based on statistics whose limits depend on the estimated eigenvalues, a fast method of calculating replications of these limits must be employed. The issues of bandwidth and kernel selection have been extensively studied in the econometric literature for over three decades; we cannot cite dozens, if not hundreds, of papers devoted to them. Perhaps the best known contributions are those of Andrews (1) and Andrews and Monahan (2) who introduced data-driven bandwidth selection and prewhitening. While these approaches possess optimality properties in general regression models with heteroskedastic and correlated errors, they are not optimal in all specific applications. In particular, Jönsson (28) found that the finite-sample distribution of the (scalar) KPSS test statistic can be very unstable when the Quadratic Spectral kernel (recommended by Andrews (1)) is used and/or a prewhitening filter is applied. He recommends the Bartlett kernel. An elaboration on the finite sample properties of the KPSS test with many relevant references can be found in Jönsson (29). This chapter focuses on the derivation and large sample theory for the stationarity tests for functional time series; we cannot present here a comprehensive and conclusive study of the finite sample properties, which are still being investigated even for scalar time series. We however wish to offer some practical guidance and report approaches which worked well for the data-generating processes we considered.

Politis (2003, 2011) argues that the flat top kernel

$$K(t) = \begin{cases} 1, & 0 \leq t < 0.1 \\ 1.1 - |t|, & 0.1 \leq t < 1.1 \\ 0, & |t| \geq 1.1 \end{cases} \tag{3.50}$$

has better properties than the Bartlett or the Parzen kernels. In our empirical work, we used kernel (3.50). Our simulations showed that $h = N^{1/2}$ is satisfactory for our hypothesis testing problem when the observations are independent or weakly dependent (functional autoregressive processes). The empirical sizes and power functions change little if h is taken ± 5 lags smaller or larger. We note that the optimal rates derived in Andrews (1) do not apply to kernel (3.50) because this piecewise function does not satisfy the regularity conditions assumed by Andrews (1). It can be shown that the optimal rates for Bartlett and Parzen kernels remain the same in the functional case, but the multiplicative constants depend in a very complex way on the high-order moments of the functions, and the arguments Andrews (1) used to approximate them cannot be readily extended.

Once the kernel and the bandwidth have been selected, the eigenvalues $\hat{\lambda}_i$ can be computed. This allows us to compute the normalized statistics $T_N^0(d)$ and $M_N^0(d)$ and use the tests based on the asymptotic distribution of their limits. The critical values can be computed by using the expansions analogous to (3.51) or (3.52) (without the $\hat{\lambda}_i$). Alternatively, since these limits do not depend on the distribution of the data, the critical values can be obtained by calculating a large number of replications of $T_N^0(d)$ and $M_N^0(d)$ for any specific functional time series. We used iid Brownian motions, and we refer to the tests which use the critical values so obtained as $T_N^0(d)(AM)$ and $M_N^0(d)(AM)$ (Alternative method). This method is extremely computationally intensive, if its performance is to be assessed by simulations; we needed almost two months of run time on the University of Utah Supercomputer (as of June 2013) to obtain the empirical rejection rates for $T_N^0(d)(AM)$ and $M_N^0(d)(AM)$ for samples of size 100 and 250 and values of d between 1 and 10.

The limits of statistics T_N and T_N^* must be approximated by the MC distribution of $\sum_{i=1}^d \hat{\lambda}_i \int B_i^2(x)dx$, and one must proceed analogously for M_N and M_N^*. Using the expansions discussed in Shorack and Wellner (49), pp. 210–211, we use the approximations

$$\hat{T}_{d,J} = \sum_{i=1}^{d} \hat{\lambda}_i \sum_{j=1}^{J} \frac{Z_j^2}{j^2\pi^2}, \tag{3.51}$$

and

$$\hat{M}_{d,J} = \sum_{i=1}^{d} \hat{\lambda}_i \sum_{j=1}^{J} \frac{Z_{2j-1}^2 + Z_{2j}^2}{4j^2\pi^2}, \tag{3.52}$$

where $\{Z_j\}_{j=1}^{\infty}$ are iid standard normal random variables. For large J, the sums over j approximate the integrals of the functionals of the Brownian bridge and eliminate the need to generate its trajectories and to perform numerical integration. In our work, we used $J = 100$, and one thousand replications to obtain MC distributions.

To select d, we use the usual "cumulative variance" approach recommended by Ramsay and Silverman (46) and Horváth and Kokoszka (25); d is chosen so that roughly $v\%$ of the sample variance is explained by the first d principal components. In our implementation, we estimated the total of 49 largest eigenvalues (the largest number under which the estimation is numerically stable), and used $d = d_v$ such that

$$\frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{d_v}}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{49}} \approx v.$$

A general recommendation is to use v equal to about 90%, but we report results for $v = .85, .90, .95$, to see how the performance of the tests is affected by the choice of d. This is a new aspect of the stationarity tests, which reflects the infinite dimensional structure of the functional data, and which is absent in tests for scalar or vector time series.

3.4.2 Empirical size and power

We first compare the empirical size of the tests implemented as described above. We consider two data-generating processes (DGPs): 1) iid copies of the Brownian motion (BM),

2) the functional AR process of order 1 (FAR(1)). There are a large number of stationary functional time series that could be considered. In our small simulation study, the focus on the BM is motivated by the application to cumulative intraday returns considered in Section 3.5; they approximately look like realizations of the BM; see Figure 3.1. The FAR(1), with Brownian motion innovations, is used to generate temporal dependence: the tests should have correct size for general stationary functional time series, not just for iid functions. The FAR(1) process is defined by the equation

$$X_i(t) = \int_0^1 \psi(t,u) X_{i-1}(u)du + W_i(t), \quad 0 \le t \le 1, \tag{3.53}$$

where the W_i are independent Brownian motions on $[0,1]$, and ψ is a kernel whose operator norm is not too large. The precise condition is somewhat technical; see Bosq (9) or Chapter 13 of Horváth and Kokoszka (25). A sufficient condition for a stationary solution to equation (3.53) to exist is that the Hilbert–Schmidt norm of ψ be less than 1. We work with the kernel

$$\psi(t,s) = c \exp\left(\frac{t^2+s^2}{2}\right)$$

with $c = .3416$ so that the Hilbert–Schmidt norm of ψ is approximately 0.5.

We consider functional time series of length $N = 100$ and $N = 250$. Each DGP is simulated one thousand times, and the percentage of rejections of the null hypothesis is reported at the significance levels of 10 and 5%. The empirical sizes are reported in Table 3.1, which leads to the following conclusions:

1. The tests $T_N^0(AM)$ and $T_N^0(AM)$ have reasonably good empirical size, which does not depend on v. Note that we used the BM processes to obtain the critical values, so it is not surprising that we observe good results when using BM as the DGP. However, the observations of the FAR(1) series are no longer BMs.

2. If the limit distribution is used to calculate the critical values, the tests based on the MC distributions (statistics T_N, M_N, T_N^*, M_N^*) are less sensitive to the choice of the cumulative variance v.

3. The tests based on M_N and M_N^* are generally too conservative at the 5% level.

4. Even though statistic T_N^* is too conservative at the 5% level in case of the FAR(1) model, it achieves a reasonable balance of empirical size at the 10 and 5% levels.

5. If the temporal dependence is not too strong, we recommend statistics T_N^* with $v =$ 90%.

We now turn to the investigation of the empirical power. The number of DGPs that could be considered under the alternative of nonstationarity is enormous. In our simulation study, we consider merely two examples intended to illustrate the theory developed in Section 3.3. Under the change point alternative, $H_{A,1}$, the DGP is

$$X_i(t) = \begin{cases} B_i(t) & \text{if } i < \lfloor N/2 \rfloor \\ B_i(t) + \delta(t) & \text{if } i \geq \lfloor N/2 \rfloor, \end{cases}$$

where the B_i are iid Brownian bridges, and $\delta(t) = 2t(1-t)$, so that the change in the mean function is comparable to the typical size of the Brownian bridge. Under the $I(1)$ alternative, $H_{A,2}$, we consider the integrated functional sequence defined by

$$X_i(t) = X_{i-1}(t) + B_i(t), \quad 1 \leq i \leq N,$$

where $X_0(t) = B_0(t)$, and $\{B_i(t)\}_{i=0}^{\infty}$ are iid Brownian Bridges. Again, each data-generating process is simulated 1000 times and the rejection rate of H_0 is reported when the significance level is 10% and 5%. Table 3.2 shows the results of these simulations. The following conclusions can be reached:

1. Under the change point alternative, the T statistics have higher power than the M statistics. This is in perfect agreement with Theorems 3.3.1 and 3.3.2, which show that the leading terms of the T statistics are four times larger than those of the corresponding M statistics.

2. The same observation remains true under the integrated alternative, and again it agrees with the theoretical rates obtained in Theorems 3.3.3 and 3.3.4. The multiplicative constants of leading terms of the T statistics are equal to second moments and those of the M statistics to corresponding variances.

3. As for empirical size, the T statistics are not sensitive to the choice of v.

4. The test based on T_N^* has slightly lower power than those based on T_N^0 and T_N, but this is because the latter two tests have slightly inflated sizes. Our overall recommendation remains to use T_N^* with $v = 0.90$. However, if very high power is of central importance, and computational time not a big concern, the method $T_N^0(AM)$ might be superior.

3.5 Application to intraday price curves

Some of the most natural and obvious functional data are intraday price curves; five such functions are shown in Figure 3.2. Not much quantitative research has however focused on the analysis of the information contained in the *shapes* of such curves, even though they very closely reflect the reactions and expectations of intraday investors. Extensive research has focused on scalar or vector summary statistics derived from intraday data, including realized volatility and noise variance estimation; see Barndorff-Nielsen and Shephard (6) and Wang and Zou (51), among many others. Several papers have however considered the shapes of suitably defined price or volatility curves; see Gabrys et al. (15), Müller et al. (38), Gabrys et al. (16), Kokoszka and Reimherr (32), and Kokoszka et al. (33). This chapter focuses on statistical methodology and the underlying theory, and we cannot include a comprehensive empirical study of functional aspects of intraday price data. We merely show that the application of our tests leads to meaningful and useful insights.

Suppose $P_n(t_j), n = 1, \dots, N, j = 1, \dots, m$, is the price of a financial asset at time t_j on day n. Figure 3.2 shows five functional data objects constructed from the 1-minute average price of Disney stock interpolated by B-splines. In this case, the number of points t_j used to construct each object is $m = 390$. Each object is viewed as a continuous curve, making these data an excellent candidate for functional data analysis. As daily closing prices form a nonstationary scalar time series, we would expect the daily price curves to form a nonstationary functional time series. When our tests are applied to sufficiently long periods of time, they indeed always reject the null hypothesis of stationarity. For shorter periods of time, H_0 is sometimes rejected and sometimes is not, most likely due to reduced power. To illustrate, Figure 3.3 displays the P–values for the test based on T_N applied to consecutive nonoverlapping segments of length N in the time period from 04/09/1997 to 04/02/2007, which comprises 2,510 trading days. This means that there are 50 segments of length $N = 50$, 25 segments of length $N = 100$, and 10 segments of length $N = 250$. If $N = 250$, H_0 is always rejected. We obtained very similar results for the other T statistics. When the M statistics are used, the rejection rates are marginally lower, but overall commensurate with those for the T statistics. We also applied the tests to several other stocks over the same period, including Chevron, Bank of America, Microsoft, IBM, McDonalds, and Walmart, and obtained nearly identical results. The results are also very similar for gold futures. The price of gold increased five fold between 2001 and 2011, with an almost linear trend. For segments of length $N = 100$, the null is sometimes not rejected if the curves do not show a clear increasing tendency over that period, but otherwise we obtained strong rejections.

In order to fit stationary functional time series models to intraday price curves; a suitable transformation should be applied. Gabrys et al. (15) put forward the following definition.

Definition 3.5.1. *Suppose $P_n(t_j), n = 1, \dots, N, j = 1, \dots, m$, is the price of a financial* *asset at time t_j on day n. The functions*

$$R_n(t_j) = 100[\ln P_n(t_j) - \ln P_n(t_1)], \qquad j = 1, 2, \ldots, m, \qquad n = 1, \ldots, N,$$

are called the cumulative intraday returns *(CIDRs).*

The idea behind Definition 3.5.1 is very simple. If the return from the start of a trading day until its close remains within the 5% range, $R_n(t_j)$ is practically equal to the simple return $[P_n(t_j) - P_n(t_1)]/P_n(t_1)$. Since $P_n(t_1)$ is fixed for every trading day, the $R_n(t_j)$ have practically the same *shape* as the price curves; see Figure 3.1. However, since they always start from zero, level stationarity is enforced. The division by $P_n(t_1)$ helps reduce the scale inflation. It can thus be hoped that the CIDRs will form a stationary functional time series, which will be amenable to the statistical analysis of the shapes of the intraday price curves. We note that the CIDRs are not readily comparable to daily returns because they do not include the overnight price change. They are designed to statistically analyze the evolution of the intraday shapes of an asset.

We wish to verify our conjecture of the stationarity of the CIDRs by application of our tests of stationarity. If the conjecture is true, the expectation is that the P-values will be roughly uniformly distributed on $(0, 1)$. Figure 3.4 shows results of the test using T_N when applied to sequential segments of the CIDR curves of the Disney stock. We see that the P-values appear to be uniformly distributed, which is consistent with the stationarity of the CIDRs. Again, the results for the other eight stocks are very similar.

3.6 Proofs of the results of Section 3.2

The proof of Theorem 3.2.1 is based on an approximation developed in Berkes et al. (8) (Theorem 3.6.1 below). Define

$$\Gamma(x,t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x) \varphi_i(t), \tag{3.54}$$

where W_i are independent and identically distributed Wiener processes (standard Brownian motions). Clearly, $\Gamma(x,t)$ is Gaussian with zero mean and $E\Gamma(x,t)\Gamma(y,s) = \min(x,y)C(t,s)$.

Theorem 3.6.1. *If assumptions* (3.1)–(3.4) *hold, then*

$$\sum_{\ell=1}^{\infty} \lambda_\ell < \infty \tag{3.55}$$

and for every N we can define a sequence of Gaussian processes $\Gamma_N(x,t)$ such that

$$\{\Gamma_N(x,t), 0 \le x,t \le 1\} \stackrel{\mathcal{D}}{=} \{\Gamma(x,t), 0 \le x,t \le 1\}$$

and

$$\sup_{0\le x\le 1} \int (V_N(x,t) - \Gamma_N(x,t))^2 dt = o_P(1),$$

where

$$V_N(x,t) = \frac{1}{N^{1/2}} \sum_{i=1}^{\lfloor Nx \rfloor} \eta_i(t).$$

(It follows immediately from (3.55) *that* $\sup_{0\le x\le 1} \int \Gamma^2(x,t)dt < \infty$ *a.s.)*

Proof of Theorem 3.2.1 Let

$$V_N^0(x,t) = V_N(x,t) - xV_N(1,t).$$

Under H_0

$$Z_N(x,t) = V_N^0(x,t) + \mu(t) \left[\frac{\lfloor Nx \rfloor - Nx}{N^{1/2}}\right]$$

and since $\mu \in L^2$, we get

$$\sup_{0\le x\le 1} ||Z_N(x,t) - V_N^0(x,t)|| \le \frac{1}{N^{1/2}} ||\mu||.$$

Hence

$$T_N = \iint (V_N^0(x,t))^2 dtdx + o_P(1)$$

and

$$M_N = \iint \left(V_N^0(x,t) - \int V_N^0(y,t)dy\right)^2 dxdt + o_P(1).$$

Applying Theorem 3.6.1, we get immediately that

$$T_N \quad \xrightarrow{\mathcal{D}} \quad \iint (\Gamma^0(x,t))^2 dxdt$$

and

$$M_N \quad \xrightarrow{\mathcal{D}} \quad \iint \left(\Gamma^0(x,t) - \int \Gamma^0(y,t)dy\right)^2 dxdt,$$

where

$$\Gamma^0(x,t) = \Gamma(x,t) - x\Gamma(1,t).$$

We also note that by the definition of $\Gamma(x,t)$ in (3.54), we have

$$\Gamma^0(x,t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} B_i(x) \varphi_i(t), \tag{3.56}$$

where B_i are independent and identically distributed Brownian bridges. Using the fact that

$\{\varphi_i(t), 0 \le t \le 1\}_{i=1}^{\infty}$ is an orthonormal system, one can easily verify that

$$\iint (\Gamma^0(x,t))^2 dxdt = \sum_{i=1}^{\infty} \lambda_i \int B_i^2(x)dx$$

and

$$\iint \left(\Gamma^0(x,t) - \int \Gamma^0(y,t)dy\right)^2 dtdx = \sum_{i=1}^{\infty} \lambda_i \int \left(B_i(x) - \int B_i(y)dy\right)^2 dx.$$

The following lemma is an immediate consequence of the results in Section 2.7 of Horváth et al. (27), or of Dunford and Schwartz (13).

Lemma 3.6.1. *If assumptions* (3.1)–(3.4), (3.13)–(3.16), (3.17), *and* H_0 *hold, then*

$$\max_{1\le i\le d} |\hat{\lambda}_i - \lambda_i| = o_P(1) \quad \text{and} \quad \max_{1\le i\le d} ||\hat{\varphi}_i - \hat{c}_i\varphi_i|| = o_P(1),$$

where $\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_d$ *are unobservable random signs defined as* $\hat{c}_i = \mathrm{sign}(\langle \hat{\varphi}_i, \varphi_i\rangle)$.

Proof of Theorem 3.2.2 It follows from Theorem 3.6.1 that

$$\sup_{0\le x\le 1} |\langle S_N(x,\cdot) - \Gamma_N^0(x,\cdot), \varphi_i\rangle| \le \sup_{0\le x\le 1} ||S_N(x,\cdot) - \Gamma_N^0(x,\cdot)|| = o_P(1)$$

and by Lemma 3.6.1, we get

$$\sup_{0\le x\le 1} |\langle \Gamma_N^0(x,\cdot), \hat{\varphi}_i - \hat{c}_i\varphi_i\rangle| \le \sup_{0\le x\le 1} ||\Gamma_N^0(x,\cdot)|| \, ||\hat{\varphi}_i - \hat{c}_i\varphi_i|| = o_P(1).$$

It is immediate from (3.56) that for all N

$$\left\{\langle \Gamma_N^0(x,\cdot), \varphi_i\rangle, 0 \le x \le 1, 1 \le i \le d\right\} \stackrel{\mathcal{D}}{=} \left\{\lambda_i^{1/2} B_i(x), 0 \le x \le 1, 1 \le i \le d\right\},$$

where $B_1, B_2, \ldots, B_d$ are independent Brownian bridges. Thus we obtain that

$$\sum_{i=1}^{d} \frac{1}{\hat{\lambda}_i} \langle \Gamma_N^0(x,\cdot), \hat{c}_i\varphi_i\rangle^2 \quad \xrightarrow{\mathcal{D}[0,1]} \quad \sum_{i=1}^{d} B_i^2(x). \tag{3.57}$$

The weak convergence in (3.57) now implies (3.18). The same arguments can be used to prove (3.19)–(3.21). □

3.7 Proofs of the results of Section 3.3

Proof of Theorem 3.3.1 First we introduce the function

$$\delta_N(x,t) = \mu(t)\{\lfloor Nx\rfloor - Nx\} + \delta(t)\{(\lfloor Nx\rfloor - k^*)I\{k^* \le \lfloor Nx\rfloor\} - x(N - k^*)\}.$$

Under $H_{A,1}$, we can write

$$Z_N(x,t) = V_N^0(x,t) + N^{-1/2}\delta_N(x,t) \tag{3.58}$$

and therefore,

$$\begin{aligned} T_N &= \iint Z_N^2(x,t)dtdx \\ &= \iint (V_N^0(x,t))^2 dtdx + \frac{2}{N^{1/2}} \iint V_N^0(x,t)\delta_N(x,t)dxdt \\ &\quad + \frac{1}{N}\iint \delta_N^2(x,t)dtdx. \end{aligned} \tag{3.59}$$

It follows from Theorem 3.6.1 that

$$\iint (V_N^0(x,t))^2 dtdx = O_P(1). \tag{3.60}$$

It is easy to check that

$$\sup_{0\le x\le 1}\left|\left|\frac{1}{N}\delta_N(x,t) - \delta_\tau(x,t)\right|\right| = O\left(\frac{1}{N}\right), \tag{3.61}$$

where $\delta_\tau(x,t)$ is defined in (3.23). Thus applying Theorem 3.6.1, we conclude that

$$\frac{1}{N}\iint V_N^0(x,t)\delta_N(x,t)dxdt \ \xrightarrow{\mathcal{D}}\ \iint \Gamma^0(x,t)\delta_\tau(x,t)dtdx. \tag{3.62}$$

Also,

$$\begin{aligned} &\frac{1}{N}\iint \delta_N^2(x,t)dtdx \\ &\quad = N\int \delta^2(t)dt\left\{\int_0^\tau x^2(1-\tau)^2dx + \int_\tau^1 (1-x)^2\tau^2dx\right\} + O(1). \end{aligned} \tag{3.63}$$

Now (3.24) is an immediate consequence of (3.59)–(3.63).

The second part of Theorem 3.3.1 is proven analogously.

3.7.1 Variances of the limits in Theorem 3.3.1

The next lemma is used to show that the variances of the limits in Theorem 3.3.1 are strictly positive.

Lemma 3.7.1. *Let Θ be a L^2 valued Gaussian process such that $E\Theta(t) = 0$ and $E\Theta(t)\Theta(s)$ is a strictly positive definite function on $[0,1]^2$. Let $g \in L^2$. Then var$(\int \Theta(t)g(t)dt) = 0$ if and only if $g = 0$ a.e.*

Proof. By the Karhunen–Loéve expansion and the assumption that $E\Theta(t)\Theta(s)$ is strictly positive definite, we may write

$$\Theta(t) = \sum_{\ell=1}^{\infty} \rho_\ell N_\ell \phi_\ell(t), \quad 0 \le t \le 1,$$

where $\{N_i\}_{i=1}^{\infty}$ are iid standard normal random variables, $\{\phi_i(t)\}_{i=1}^{\infty}$ form an orthonormal basis, and $\rho_i > 0$ for all $i \ge 1$. It follows by a simple calculation that

$$\int \Theta(t) g(t) dt = \sum_{\ell=1}^{\infty} \rho_\ell N_\ell \langle \phi_\ell, g \rangle,$$

and hence

$$\operatorname{var}\left(\int \Theta(t) g(t) dt \right) = \sum_{\ell=1}^{\infty} \rho_\ell^2 \langle \phi_\ell, g \rangle^2.$$

Since $\sum_{\ell=1}^{\infty} \rho_\ell^2 \langle \phi_\ell, g \rangle^2 = 0$ if and only if $g = 0$ *a.e.*, the result follows.

$\square$

It is easy to see that $\iint \Gamma^0(x,t)\delta_\tau(x,t)dtdx$ is a normal random variable with zero mean. Its variance is thus equal to

$$\begin{aligned} &E\left(\iint \Gamma^0(x,t)\delta_\tau(x,t)dtdx \right)^2 \qquad (3.64) \\ &= \iiiint C(t,s)\delta_\tau(x,t)\delta_\tau(y,s)(\min(x,y) - xy)dtdsdxdy \\ &= \left(\iint C(t,s)\delta(t)\delta(s)dtds \right) \left(\iint \bar{\delta}_\tau(x)\bar{\delta}_\tau(y)(\min(x,y) - xy)dxdy \right), \end{aligned}$$

where $\bar{\delta}_\tau(x)$ is defined in (3.39). Similarly to (3.24), the limit in (3.25) is normally distributed with zero mean and variance equal to

$$\begin{aligned} &E\left[\iint \left(\Gamma^0(x,t) - \int \Gamma^0(y,t)dy \right) \left(\delta_\tau(x,t) - \int \delta_\tau(y,t)dy \right) dtdx \right]^2 \\ &= \left(\iint C(t,s)\delta(t)\delta(s)dtds \right) \left(\iint \bar{\delta}_\tau(x)\bar{\delta}_\tau(y) \Big[\min(x,y) - xy - \int (\min(y,z) - yz)dz \right. \\ &\qquad\qquad \left. - \int (\min(x,z) - xz)dz + \iint (\min(z,z') - zz')dzdz' \Big] dxdy \right) \\ &= \left(\iint C(t,s)\delta(t)\delta(s)dtds \right) \left(\iint \bar{\delta}_\tau(x)\bar{\delta}_\tau(y) \Big[\min(x,y) - xy - \frac{y(1-y)}{2} \right. \\ &\qquad\qquad\qquad\qquad \left. - \frac{x(1-x)}{2} + \frac{1}{12} \Big] dxdy \right). \end{aligned}$$

If the bivariate function $C(t,s)$ is strictly positive definite, then $\iint C(t,s)\delta(t)\delta(s)dtds > 0$ if $\delta(t)$ is not the 0 function in L^2. Observing that $\iint \bar{\delta}_\tau(x)\bar{\delta}_\tau(y)(\min(x,y) - xy)dxdy =$

var($\int B(x)\bar{\delta}_\tau(x)$), where B is a Brownian bridge, the positivity of (3.64) follows by Lemma 3.7.1 since $\bar{\delta}_\tau(x)$ is not the zero function and the covariance function of the Brownian bridge is strictly positive definite. A similar application of Lemma 3.7.1 yields that

$$\iint \bar{\delta}_\tau(x)\bar{\delta}_\tau(y)\Big[\min(x,y) - xy - \frac{y(1-y)}{2} - \frac{x(1-x)}{2} + \frac{1}{12}\Big]dxdy > 0.$$

Lemma 3.7.2. *If assumptions* (3.1)–(3.4), (3.13)–(3.16), (3.22), *and* $H_{A,1}$ *hold, then*

$$\left\|\hat{C}_N(t,s) - \left(2\tau(1-\tau)\delta(t)\delta(s)\sum_{i=1}^{N} K(i/h) + \bar{C}_N(t,s)\right)\right\| = O_P(h/N^{1/2}),$$

where

$$\bar{C}_N(t,s) = \bar{\gamma}_0(t,s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\{\bar{\gamma}_i(t,s) + \bar{\gamma}_i(s,t)\} \tag{3.65}$$

with

$$\bar{\gamma}_i(t,s) = \frac{1}{N}\sum_{j=i+1}^{N}\left(\eta_j(t) - \bar{\eta}_N(t)\right)\left(\eta_{j-i}(s) - \bar{\eta}_N(s)\right), \quad 0 \le i \le N-1.$$

Proof. First we write with $\mu_i(t) = EX_i(t)$ and observe that

$$\begin{aligned}
\hat{\gamma}_i(t,s) &= \frac{1}{N}\sum_{j=i+1}^{N}\left(\eta_j(t) - \bar{\eta}_N(t) - [\bar{\mu}_N(t) - \mu_i(t)]\right)\left(\eta_{j-i}(s) - \bar{\eta}_N(s) - [\bar{\mu}_N(s) - \mu_{j-i}(s)]\right)\\
&= \frac{1}{N}\sum_{j=i+1}^{N}\left(\eta_j(t) - \bar{\eta}_N(t)\right)\left(\eta_{j-i}(s) - \bar{\eta}_N(s)\right)\\
&\quad + \frac{1}{N}\sum_{j=i+1}^{N}(\eta_j(t) - \bar{\eta}_N(t))(\mu_{j-i}(s) - \bar{\mu}_N(s))\\
&\quad + \frac{1}{N}\sum_{j=i+1}^{N}(\mu_j(t) - \bar{\mu}_N(t))(\eta_{j-i}(s) - \bar{\eta}_N(s))\\
&\quad + \frac{1}{N}\sum_{j=i+1}^{N}(\mu_j(t) - \bar{\mu}_N(t))(\mu_{j-i}(s) - \bar{\mu}_N(s))\\
&= \bar{\gamma}_i(t,s) + \hat{\gamma}_i^{(1)}(t,s) + \hat{\gamma}_i^{(2)}(t,s) + \hat{\gamma}_i^{(3)}(t,s)
\end{aligned}$$

with

$$\bar{\eta}_N(t) = \frac{1}{N}\sum_{\ell=1}^{N}\eta_\ell \qquad \text{and} \qquad \bar{\mu}_N(t) = \mu(t) + \frac{N - \lfloor N\tau\rfloor}{N}\delta(t).$$

By the triangle inequality, we have

$$\left|\left|\hat{\gamma}_0^{(1)}(t,s)+\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)(\hat{\gamma}_i^{(1)}(t,s)+\hat{\gamma}_i^{(1)}(s,t))\right|\right|$$
$$\leq \left|\left|\hat{\gamma}_0^{(1)}(t,s)\right|\right|+\left|\left|\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\hat{\gamma}_i^{(1)}(t,s)\right|\right|+\left|\left|\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\hat{\gamma}_i^{(1)}(s,t)\right|\right|.$$

Using Theorem 3.6.1, we get

$$\left|\left|\hat{\gamma}_0^{(1)}(t,s)\right|\right| = O_P(N^{-1/2}).$$

Using again the triangle inequality, we obtain that

$$E\left|\left|\sum_{i=1}^{N-1}K\left(i/h\right)\hat{\gamma}_i^{(1)}(t,s)\right|\right| \leq \sum_{i=1}^{N-1}K\left(i/h\right)E||\hat{\gamma}_i^{(1)}(t,s)||. \tag{3.66}$$

Furthermore, by an application of the Cauchy–Schwarz inequality,

$$E||\hat{\gamma}_i^{(1)}(t,s)||\leq \left|\left|\frac{1}{N}\sum_{j=i+1}^{N}(\mu_{j-i}(s)-\bar{\mu}_N(s))\right|\right| E\left|\left|\frac{1}{N}\sum_{j=i+1}^{N}(\eta_j(t)-\bar{\eta}_N(t))\right|\right|.$$

It is clear that

$$\max_{1\leq i\leq N}\left|\left|\frac{1}{N}\sum_{j=i+1}^{N}(\mu_{j-i}(s)-\bar{\mu}_N(s))\right|\right| = O(1),$$

and by Berkes et al. (2013),

$$\max_{1\leq i\leq N}E\left|\left|\frac{1}{N}\sum_{j=i+1}^{N}(\eta_j(t)-\bar{\eta}_N(t))\right|\right| = O(N^{-1/2}).$$

Combining these bounds with (3.66) and assumptions (3.13)–(3.15) gives

$$E\left|\left|\sum_{i=1}^{N-1}K(i/h)\hat{\gamma}_i^{(1)}(t,s)\right|\right| = O(h/N^{1/2}),$$

and hence by Markov's inequality,

$$\left|\left|\sum_{i=1}^{N-1}K(i/h)\hat{\gamma_i^{(1)}}(t,s)\right|\right| = O_P(h/N^{1/2}).$$

Thus we conclude

$$\left|\left|\hat{\gamma}_0^{(1)}(t,s)+\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)(\hat{\gamma}_i^{(1)}(t,s)+\hat{\gamma}_i^{(1)}(s,t))\right|\right| = O_P(h/N^{1/2}). \tag{3.67}$$

Similarly to (3.67), we have

$$\left|\left|\hat{\gamma}_0^{(2)}(t,s)+\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)(\hat{\gamma}_i^{(2)}(t,s)+\hat{\gamma}_i^{(2)}(s,t))\right|\right|=O_P(h/N^{1/2}). \tag{3.68}$$

Using the definition of $\bar{\mu}_N(t)$ and $H_{A,1}$, we obtain that

$$\max_{0\le i\le h}||\hat{\gamma}_0^{(3)}(t,s)-\tau(1-\tau)\delta(t)\delta(s)||=O(h/N). \tag{3.69}$$

The lemma now follows from (3.67)–(3.69). □

Proof of Theorem 3.3.2 The proof of Theorem 3.3.2 is based on the asymptotic properties of $\hat{C}_N$ under $H_{A,1}$. It follows from Lemma 3.7.2 that (3.26) and (3.27) hold assuming only (3.16). We write by (3.58)

$$\langle Z_N(x,\cdot),\hat{\varphi}_1\rangle^2=\langle V_N^0(x,\cdot),\hat{\varphi}_1\rangle^2+N^{-1}\langle\delta_N(x,\cdot),\hat{\varphi}_1\rangle^2+2\langle V_N^0(x,\cdot),\hat{\varphi}_1\rangle N^{-1/2}\langle\delta_N(x,\cdot),\hat{\varphi}_1\rangle.$$

Combining Theorem 3.6.1 with the Cauchy–Schwarz inequality, we get

$$\sup_{0\le x\le 1}|\langle V_N^0(x,\cdot),\hat{\varphi}_1\rangle|\le\sup_{0\le x\le 1}||V_N^0(x,\cdot)||=O_P(1).$$

Using (3.61), we conclude

$$\int N^{-1}\langle\delta_N(x,\cdot),\hat{\varphi}_1\rangle^2dx=\frac{N}{3}\tau^2(1-\tau)^2\langle\delta,\hat{\varphi}_1\rangle^2(1+O_P(1/N)).$$

Theorem 3.6.1 and (3.27) yield

$$\begin{aligned}
&N^{1/2}\int\langle V_N^0(x,\cdot),\hat{\varphi}_1\rangle\langle\delta_N(x,\cdot),\hat{\varphi}_1\rangle dx\\
&\quad\xrightarrow{\mathcal{D}}\frac{1}{||\delta||}\int\langle\Gamma_0(x,\cdot),\delta\rangle\langle\delta_\tau(x,\cdot),\delta\rangle dx\\
&\quad=\int\left\{\int\Gamma^0(x,t)\delta(t)dt\right\}[(x-\tau)I\{x\ge\tau\}-x(1-\tau)]dx\\
&\quad=\iint\Gamma^0(x,t)\delta_\tau(x,t)dxdt.
\end{aligned}$$

This completes the proof of (3.29). It follows from (3.29) that

$$\frac{\hat{\lambda}_1}{N^{1/2}}\left\{T_N^0(1)-\frac{N}{3\hat{\lambda}_1}\tau^2(1-\tau)^2\langle\delta,\hat{\varphi}_1\rangle^2\right\}\quad\xrightarrow{\mathcal{D}}\quad 2\iint\Gamma^0(x,t)\delta_\tau(x,t)dxdt,$$

and therefore, (3.29) implies (3.31). Similar arguments prove (3.30) and (3.32).

If in addition we assume that $h/N^{1/2}\to 0$ as $N\to\infty$, then by Lemma 3.7.2 and Dunford and Schwartz (13), we have (3.26), (3.27), and for every fixed $i\ge 2$,

$$\hat{\lambda}_i\quad\xrightarrow{P}\quad\bar{\lambda}_i, \tag{3.70}$$

where $\bar\lambda_2 \geq \bar\lambda_3 \geq \ldots \geq 0$ (different from the $\lambda_i, i \geq 2$),

$$||\hat\varphi_i(t) - \hat c_i \bar\varphi_i|| = o_P(1), \quad i \geq 2, \tag{3.71}$$

with some functions $\bar\varphi_2, \bar\varphi_3, \ldots$, where $\hat c_i = \mathrm{sign}(\langle \hat\varphi_i, \bar\varphi_i\rangle)$. (Of course, $\bar\varphi_i$ is only defined if $\hat\lambda_i > 0$.) Using again (3.58) with Theorem 3.6.1 and (3.71), we obtain that

$$\int \langle Z_N(x,\cdot), \hat\varphi_i\rangle^2 dx = \frac{N}{3}\tau^2(1-\tau)^2\langle \delta, \hat\varphi_i\rangle^2 + O_P(N^{1/2}).$$

Since δ and $\bar\varphi_i$ are orthogonal for all $i \geq 2$, (3.71) implies $\langle \delta, \hat\varphi_i\rangle = o_P(1)$. Hence (3.33) follows from (3.29). The results in (3.34)–(3.36) can be established similarly so the proofs are omitted.

Proof of Remark 3.1. Let

$$\beta_N(x,t) = \mu(t)\{\lfloor Nx\rfloor - Nx\} + \delta_N^*(t)\{(\lfloor Nx\rfloor - k^*)I\{k^* \leq \lfloor Nx\rfloor\} - x(N-k^*)\}.$$

Using (3.59) with $\delta_N(x,t)$ replaced with $\beta_N(x,t)$ and Theorem 3.6.1, we get

$$\begin{aligned} T_N - \frac{1}{N}||\beta_N||^2 = &\iint (\Gamma_N^0(x,t))^2 dtdx(1+o_P(1)) \\ &+ 2N^{1/2}\iint \Gamma_N^0(x,t)\delta_N^*(t)\bar\delta_\tau(x)dtdx(1+o_P(1)). \end{aligned} \tag{3.72}$$

By the Cauchy–Schwarz inequality

$$\iint \Gamma_N^0(x,t)\delta_N^*(t)\bar\delta_\tau(x)dtdx = O_P(||\delta_N^*||). \tag{3.73}$$

Elementary arguments show that

$$\frac{1}{N}||\beta_N||^2 = ||\bar\delta_\tau||^2 N||\delta_N^*||^2(1+o(1)), \tag{3.74}$$

as $N \to \infty$. If $N^{1/2}||\delta_N^*|| \to 0$ as $N \to \infty$ then by (3.72)–(3.74), we obtain immediately that $T_N \xrightarrow{\mathcal{D}} \iint(\Gamma^0(x,t))^2 dtdx$. If $N^{1/2}||\delta_N^*|| \to \infty$, then again by (3.72)–(3.74), we see that $T_N \xrightarrow{P} \infty$. Since for every fixed N, $\iint \Gamma_N^0(x,t)\delta_N^*(t)\bar\delta_\tau(x)dtdx$ is normal with zero mean and variance $\iint(\min(x,y)-xy)\bar\delta_\tau(x)\bar\delta_\tau(y)dxdy \iint \delta_N^*(t)\delta_N^*(s)C(t,s)dtds$, hence (3.37) follows. In the case when $N^{1/2}\delta_N^* \xrightarrow{L^2} \delta^*$, it follows from (3.74) that $(1/N)||\beta_N||^2 \to \zeta = ||\bar\delta_\tau||^2||\delta^*||^2 > 0$. Now by (3.72) and the representation of Γ_N^0 in (3.56), we conclude

$$
\begin{aligned}
T_N &\stackrel{\mathcal{D}}{=} \zeta(1+o(1)) \\
&\quad + \sum_{\ell=1}^{\infty} \lambda_\ell^{1/2} \left[\lambda_\ell^{1/2} \int B_\ell^2(x)dx + \int B_\ell(x)\bar{\delta}_\tau(x)dx \int \varphi_\ell(t) N^{1/2} \delta_N^*(t)dt \right] (1+o_P(1)) \\
&\to \zeta + \sum_{\ell=1}^{\infty} \left\{ \lambda_\ell ||B_\ell||^2 + 2\lambda_\ell^{1/2} \langle B_\ell, \bar{\delta_\tau} \rangle \langle \varphi_\ell, \delta^* \rangle \right\},
\end{aligned}
$$

which completes the proof of (3.38). □

Lemma 3.7.3. *If assumptions* (3.1)–(3.4) *hold, then*

$$
\sup_{0\le x\le 1} \int \left(U_N(x,t) - \int_0^x \Gamma_N(u,t)du \right)^2 dt = o_P(1), \tag{3.75}
$$

where

$$
U_N(x,t) = \frac{1}{N^{3/2}} \sum_{k=1}^{\lfloor Nx \rfloor} \sum_{i=1}^{k} \eta_i(t),
$$

and the Gaussian processes $\Gamma_N(x,t)$ are defined in Theorem 3.6.1.

Proof. It is enough to verify that

$$
\sup_{0\le x\le 1} \int \left(U_N(x,t) - \int_0^x V_N(u,t)du \right)^2 dt = \sup_{0\le x\le 1} \left|\left| U_N(x,\cdot) - \int_0^x V_N(u,\cdot) \right|\right|^2 = o_P(1)
$$

and

$$
\sup_{0\le x\le 1} \int \left(\int_0^x \{V_N(u,t) - \Gamma_N(u,t)\} du \right)^2 dt = o_P(1).
$$

Elementary arguments yield

$$
\left| U_N(x,t) - \int_0^x V_N(u,t)du \right| \le \frac{1}{N^{3/2}} \left| \sum_{i=1}^{\lfloor Nx \rfloor} \eta_i(t) \right|.
$$

It follows from Theorem 3.6.1 that

$$
\sup_{0\le x\le 1} \left|\left| N^{-1/2} \sum_{i=1}^{\lfloor Nx \rfloor} \eta_i(\cdot) \right|\right| = O_P(1),
$$

and therefore,

$$
\sup_{0\le x\le 1} \left|\left| U_N(x,\cdot) - \int_0^x V_N(u,\cdot)du \right|\right| = O_P\left(\frac{1}{N}\right).
$$

Using the Cauchy–Schwarz inequality with Theorem 3.6.1, we conclude

$$
\begin{aligned}
\int \left(\int_0^x (V_N(u,t) - \Gamma_N(u,t))\, du \right)^2 dt &\le \int \int_0^x (V_N(u,t) - \Gamma_N(u,t))^2 \, dudt \\
&\le \int \int (V_N(u,t) - \Gamma_N(u,t))^2 \, dudt
\end{aligned}
$$

$$= o_P(1).$$

Now the proof of Lemma 3.7.3 is complete. $\square$

Proof of Theorem 3.3.3 First we note that under $H_{A,2}$ we have

$$\frac{1}{N^{3/2}} \sum_{k=1}^{\lfloor Nx \rfloor} X_k(t) = U_N(x,t) + \frac{\lfloor Nx \rfloor}{N^{3/2}} \mu(t). \tag{3.76}$$

Therefore,

$$\frac{1}{N} Z_N(x,t) = U_N(x,t) - xU_N(1,t) + \frac{\lfloor Nx \rfloor - xN}{N^{3/2}} \mu(t).$$

Using (3.76), we get via the Cauchy–Schwarz inequality

$$\begin{aligned}
&\left| \iint \left(\frac{1}{N} Z_N(x,t) \right)^2 dtdx - \iint (U_N(x,t) - xU_N(1,t))^2 dtdx \right| \\
&\le \iint \left\{ \frac{1}{N} Z_N(x,t) - [U_N(x,t) - xU_N(1,t)] \right\}^2 dtdx \\
&\qquad + 2 \iint \left| \frac{1}{N} Z_N(x,t) - [U_N(x,t) - xU_N(1,t)] \right| \left| U_N(x,t) - xU_N(1,t) \right| dtdx \\
&\le o_P(1) + o_P(1) \left\{ \iint (U_N(x,t) - xU_N(1,t))^2 dtdx \right\}^{1/2} \\
&= o_P(1),
\end{aligned}$$

since by Lemma 3.7.3

$$\iint (U_N(x,t) - xU_N(1,t))^2 dtdx = O_P(1).$$

It also follows from Lemma 3.7.3 that

$$\iint (U_N(x,t) - xU_N(1,t))^2 dtdx \quad \xrightarrow{\mathcal{D}} \quad \iint \Delta^2(x,t) dtdx,$$

which completes the proof of (3.41).

The proof of (3.42) is similar to that of (3.41) and therefore, the details are omitted.

Lemma 3.7.4. *Define*

$$I_N(z,t) = \int_0^z \Gamma_N(u,t) du - \int \left\{ \int_0^v \Gamma_N(u,t) du \right\} dv,$$

where the Gaussian processes $\Gamma_N(x,t)$ are defined in Theorem 3.6.1. Let

$$Q_N(t,s) = 2 \left(\int_0^c K(w) dw \right) \int_0^1 I_N(z,t) I_N(z,s) dz.$$

If assumptions (3.1)–(3.4), (3.13)–(3.16), *and* $H_{A,2}$ *hold, then*

$$\left\|\frac{1}{Nh}\hat{C}_N(t,s) - Q_N(t,s)\right\| = o_P(1).$$

Proof. Since

$$\bar{X}_N(t) = \mu(t) + \frac{1}{N}\sum_{j=1}^{N}\sum_{\ell=1}^{j}\eta_\ell(t),$$

Theorem 3.6.1 yields

$$\left\|N^{-1/2}(\bar{X}_N(t)-\mu(t)) - \int\left\{\int_0^v \Gamma_N(u,t)du\right\}dv\right\| = o_P(1),$$

resulting in

$$\begin{aligned}\max_{1\le i\le N-1}\Bigg\|\frac{1}{N}\hat{\gamma}_i(t,s) - \frac{1}{N}\sum_{j=i+1}^{N}&\left(\int_0^{j/N}\Gamma_N(u,t)du - \int\left\{\int_0^v\Gamma_N(u,t)du\right\}dv\right) \\ &\times\left(\int_0^{(j-i)/N}\Gamma_N(u,s)du - \int\left\{\int_0^v\Gamma_N(u,s)du\right\}dv\right)\Bigg\| = o_P(1).\end{aligned} \tag{3.77}$$

Next we use the almost sure continuity with $\Gamma_N(0,t)=0$ to conclude

$$\begin{aligned}\max_{1\le i\le ch}\Bigg\|\frac{1}{N}\sum_{j=i+1}^{N}&\left(\int_0^{j/N}\Gamma_N(u,t)du - \int\left\{\int_0^v\Gamma_N(u,t)du\right\}dv\right) \\ &\times\left(\int_0^{(j-i)/N}\Gamma_N(u,s)du - \int\left\{\int_0^v\Gamma_N(u,s)du\right\}dv\right) \\ &-\frac{1}{N}\sum_{j=i+1}^{N}\left(\int_0^{j/N}\Gamma_N(u,t)du - \int\left\{\int_0^v\Gamma_N(u,t)du\right\}dv\right) \\ &\quad\times\left(\int_0^{j/N}\Gamma_N(u,s)du - \int\left\{\int_0^v\Gamma_N(u,s)du\right\}dv\right)\Bigg\| = o_P(1).\end{aligned} \tag{3.78}$$

Putting together (3.77) and (3.78), we get

$$\begin{aligned}\max_{1\le i\le ch}\Bigg\|\frac{1}{N}\hat{\gamma}_i(t,s) - \int_{i/N}^{1}&\Bigg[\left(\int_0^{z}\Gamma_N(u,t)du - \int\left\{\int_0^v\Gamma_N(u,t)du\right\}dv\right) \\ &\times\left(\int_0^{z}\Gamma_N(u,s)du - \int\left\{\int_0^v\Gamma_N(u,s)du\right\}dv\right)\Bigg]dz\Bigg\| = o_P(1)\end{aligned}$$

and

$$\max_{1\le i\le ch}\Bigg\|\int_{i/N}^{1}\Bigg[\left(\int_0^z \Gamma_N(u,t)du-\int\left\{\int_0^v \Gamma_N(u,t)du\right\}dv\right)$$
$$\times\left(\int_0^z \Gamma_N(u,s)du-\int\left\{\int_0^v \Gamma_N(u,s)du\right\}dv\right)\Bigg]dz$$
$$-\int_0^{1}\Bigg[\left(\int_0^z \Gamma_N(u,t)du-\int\left\{\int_0^v \Gamma_N(u,t)du\right\}dv\right)$$
$$\times\left(\int_0^z \Gamma_N(u,s)du-\int\left\{\int_0^v \Gamma_N(u,s)du\right\}dv\right)\Bigg]dz\Bigg\|=o_P(1).$$

Since K satisfies conditions (3.14) and (3.15), the proof of Lemma 3.7.4 is complete. □

Lemma 3.7.5. *For every $N\ge 1$ we have*

$$\left\{Q_N(t,s),0\le t,s\le 1\right\}\overset{\mathcal{D}}{=}2\int_0^c K(w)dw\left\{\sum_{i,j=1}^{\infty}\lambda_i^{1/2}\lambda_j^{1/2}\varphi_i(t)\varphi_j(s)\nu_{i,j}\right\},\tag{3.79}$$

where $\lambda_1,\lambda_2,\dots,\varphi_1,\varphi_2,\dots$ are defined (3.9) *and for every $i,j\le 1$*

$$\nu_{i,j}\overset{\mathcal{D}}{=}\int\Bigg[\left\{\int_0^z W_i(u)du-\int\left(\int_0^v W_i(u)du\right)dv\right\}$$
$$\times\left\{\int_0^z W_j(u)du-\int\left(\int_0^v W_j(u)du\right)dv\right\}\Bigg]dz,$$

where $W_1,W_2,\dots$ are independent Wiener processes. Also, $Q_N(t,s)$ is a non–negative definite function for all N with probability one.

Proof. The representation in (3.79) is an immediate consequence of (3.54). It follows from (3.79) that $Q_N(t,s)$ is symmetric and $Q_N\in L^2$ with probability one. Also for any $g\in L^2$ we have

$$\iint Q_N(t,s)g(t)g(s)dtds$$
$$=\int\left(\int\sum_{i=1}^{\infty}\lambda_i^{1/2}\varphi_i(t)\left\{\int_0^z W_i(u)du-\int\left(\int_0^v W_i(u)du\right)dv\right\}g(t)dt\right)^2dz$$
$$\ge 0,$$

completing the proof. □

Proof of Theorem 3.3.4 The result follows immediately from the proofs of Lemmas 3.7.3 and 3.7.4.

Proof of Theorem 3.3.5 The result in Theorem 3.3.4 and (3.43) yield that there are processes $\Gamma_N(x,t), \Delta_N(x,t), Q_N(t,s)$ such that

$$\{\Gamma_N(x,t), \Delta_N(x,t), Q_N(t,s), 0 \le x,t,s \le 1\} \overset{\mathcal{D}}{=} \{\Gamma(x,t), \Delta(x,t), Q_N(t,s), 0 \le x,t,s \le 1\}$$

and

$$\max_{0\le x\le 1}\left|\left|\frac{1}{N}Z_N(x,t) - \Delta_N(x,t)\right|\right| = o_P(1) \quad \text{and} \quad \left|\left|\frac{1}{Nh}\hat{C}_N(t,s) - Q_N(t,s)\right|\right| = o_P(1).$$

Similarly to (3.43) we define $\lambda_{1,N*} \ge \lambda^*_{2,N} \ge \ldots$ and random functions $\varphi^*_{1,N}(t), \varphi^*_2(t), \ldots$ satisfying

$$\lambda^*_{i,N}\varphi^*_{i,N}(t) = \int Q_N(t,s)\varphi^*_{i,N}(s)ds, \;\; 1 \le i < \infty. \tag{3.80}$$

Hence

$$\max_{1\le i\le d}|\hat{\lambda}_i - \lambda^*_{i,N}| = o_P(1)$$

and

$$\max_{1\le i\le d}||\hat{\varphi}_i - \hat{c}_i\varphi^*_{i,N}|| = o_P(1),$$

where $\hat{c}_1, \hat{c}_2, \ldots$ are random signs. By construction,

$$\begin{aligned}\{\Delta_N(x,t), &Q_N(t,s), \lambda^*_{1,N}, \ldots, \lambda^*_{d,N}, (\varphi^*_{1,N}(t))^2, \ldots, (\varphi^*_{d,N}(t))^2, 0 \le x,t,s \le 1\} \\ &\overset{\mathcal{D}}{=} \{\Delta(x,t), Q_N(t,s), \lambda^*_1, \ldots, \lambda^*_d, (\varphi^*_1(t))^2, \ldots, (\varphi^*_d(t))^2, 0 \le x,t,s \le 1\},\end{aligned}$$

which completes the proof.

3.8 Bibliography

[1] D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

[2] D. W. K. Andrews and J. C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60:953–966, 1992.

[3] A. Antoniadis and T. Sapatinas. Wavelet methods for continuous time prediction using Hilbert–valued autoregressive processes. *Journal of Multivariate Analysis*, 87:133–158, 2003.

[4] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B*, 68:837–857, 2006.

[5] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37:4046–4087, 2009.

[6] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariance: High frequency based covariance, regression and correlation in financial economics. *Econometrica*, 72:885–925, 2004.

[7] I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)*, 71:927–946, 2009.

[8] I. Berkes, L. Horváth, and G. Rice. Weak invariance principles for sums of dependent random functions. *Stochastic Processes and Their Applications*, 2013. Under revision.

[9] D. Bosq. *Linear Processes in Function Spaces.* Springer, New York, 2000.

[10] R. M. de Jong, C. Amsler, and P. Schmidt. A robust version of the KPSS test based on indicators. *Journal of Econometrics*, 137:311–333, 1997.

[11] D. A. Dickey and W. A. Fuller. Distributions of the estimattors for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

[12] D. A. Dickey and W. A. Fuller. Likelihood ratio statistics for autoregressive time series with unit root. *Econometrica*, 49:1057–1074, 1981.

[13] N. Dunford and J. T. Schwartz. *Linear Operators, Parts I and II.* Wiley, New York, 1988.

[14] Y. Dwivedi and S. Subba Rao. A test for second order stationarity based on the discrete Fourier transform. *Journal of Time Series Analysis*, 32:68–91, 2011.

[15] R. Gabrys, L. Horváth, and P. Kokoszka. Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, 105:1113–1125, 2010.

[16] R. Gabrys, S. Hörmann, and P. Kokoszka. Monitoring the intraday volatility pattern. *Journal of Time Series Econometrics*, 2013, Forthcoming

[17] L. Giraitis, P. S. Kokoszka, R. Leipus, and G. Teyssière. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 112:265–294, 2003.

[18] C. W. J. Granger and M. Hatanaka. *Spectral Analysis of Economic Time Series.* Princeton University Press, 1964.

[19] U. Grenander and M. Rosenblatt. *Statistical Analysis of Stationary Time Series.* Wiley, New York, 1957.

[20] S. Hays, H. Shen, and J. Z. Huang. Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, 6:870–894, 2012.

[21] S. Hörmann and P. Kokoszka. Weakly dependent functional data. *The Annals of Statistics*, 38:1845–1884, 2010.

[22] S. Hörmann and P. Kokoszka. Functional time series. In C. R. Rao and T. Subba Rao, editors, *Time Series*, volume 30 of *Handbook of Statistics*. Elsevier, New York, 2012.

[23] S. Hörmann, L. Horváth, and R. Reeder. A functional version of the ARCH model. *Econometric Theory*, 29:267–288, 2013.

[24] S. Hörmann, L. Kidziński, and M. Hallin. Dynamic functional principal components. Technical report, Université libre de Bruxelles, 2013.

[25] L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, New York, 2012.

[26] L. Horváth, M. Hušková, and P. Kokoszka. Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, 101:352–367, 2010.

[27] L. Horváth, P. Kokoszka, and R. Reeder. Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, 75: 103–122, 2013.

[28] K. Jönsson. Finite-sample stability of the KPSS test. Working Paper 2006:23, Department of Economics, Lund University, 2006.

[29] K. Jönsson. Testing stationarity in small- and medium-sized samples when disturbances are serially correlated. *Oxford Bulletin of Economics and Statistics*, 73:669–690, 2011.

[30] V. Kargin and A. Onatski. Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99:2508–2526, 2008.

[31] J. Kiefer. K-sample analogues of the Kolmogorov-Smirnov and Cramér-v.Mises tests. *Ann. Math. Statist.*, 30:420–447, 1959.

[32] P. Kokoszka and M. Reimherr. Predictability of shapes of intraday price curves. *Econometrics Journal*, 2013. Forthcoming.

[33] P. Kokoszka, H. Miao, and X. Zhang. Functional multifactor regression for intraday price curves. Technical report, Colorado State University, 2013.

[34] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159–178, 1992.

[35] D. Lee and P. Schmidt. On the power of the KPSS test of stationarity against fractionally integrated alternatives. *Journal of Econometrics*, 73:285–302, 1996.

[36] A. W. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.

[37] T. McMurry and D. N. Politis. Resampling methods for functional data. In F. Ferraty and Y. Romain, editors, *Oxford Handbook on Statistics and FDA*. Oxford University Press, 2010.

[38] H-G. Müller, R. Sen, and U. Stadtmüller. Functional data analysis for volatility. *Journal of Econometrics*, 165:233–245, 2011.

[39] V. M. Panaretos and S. Tavakoli. Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41:568–603, 2013.

[40] V. M. Panaretos and S. Tavakoli. Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, 123:2779–2807, 2013.

[41] M. M. Pelagatti and P. K. Sen. Rank tests for short memory stationarity. *Journal of Econometrics*, 172:90–105, 2013.

[42] D. N. Politis. Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 25: 517–533, 2003.

[43] D. N. Politis. Higher-order accurate, positive semidefinite estimation of large sample covariance and spectral density matrices. *Econometric Theory*, 27:1469–4360, 2011.

[44] B. Pötscher and I. Prucha. *Dynamic Non–linear Econonometric Models. Asymptotic Theory.* Springer, New York, 1997.

[45] M. B. Priestley and T. Subba Rao. A test for non–stationarity of time–series. *Journal of the Royal Statistical Society (B)*, 31:140–149, 1969.

[46] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer, New York, 2005.

[47] S. E. Said and D. A. Dickey. Testing for unit roots in autoregressive–moving average models of unknown order. *Biometrika*, 71:599–608, 1984.

[48] X. Shao and W. B. Wu. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35:1773–1801, 2007.

[49] G. R. Shorack and J. A. Wellner. *Empirical Processes with Applications to Statistics.* Wiley, New York, 1986.

[50] T. Teräsvirta, D. Tjøstheim, and C. W. J. Granger. *Modeling Nonlinear Economic Time Series.* Advanced Texts in Econometrics. Oxford University Press, 2010.

[51] Y. Wang and J. Zou. Vast volatility matrix estimation for high–frequency financial data. *The Annals of Statistics*, 38:953–978, 2010.

[52] W. Wu. *Nonlinear System Theory: Another Look at Dependence*, volume 102 of *Proceedings of The National Academy of Sciences of the United States.* National Academy of Sciences, 2005.

[53] X. Zhang, X. Shao, K. Hayhoe, and D. Wuebbles. Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics*, 5:1765–1796, 2011.

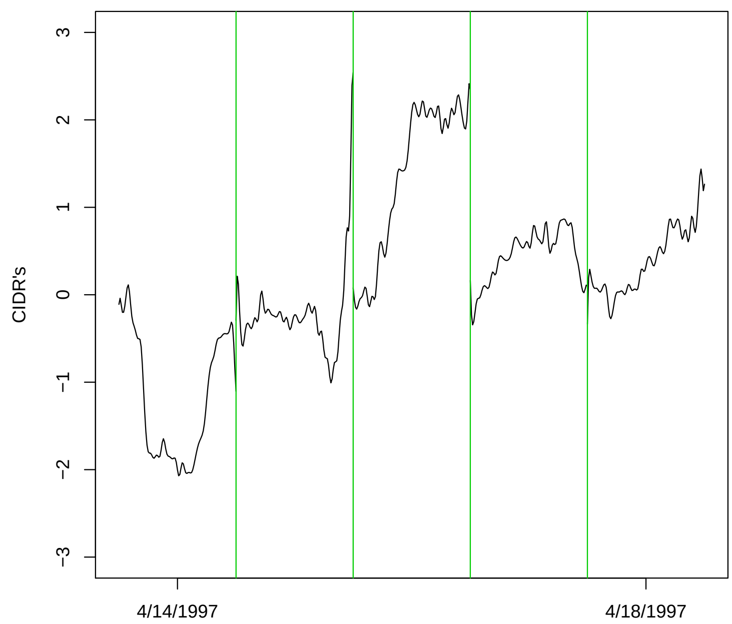

Figure 3.1. Five cumulative intraday returns constructed from the intraday prices displayed in Figure 3.2.

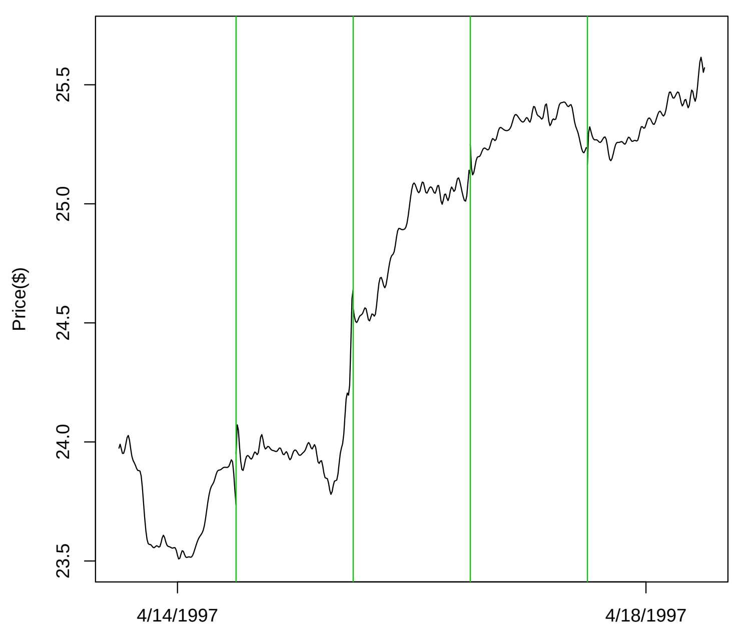

Figure 3.2. Five functional data objects constructed from the 1-minute average price of Disney stock. The vertical lines separate the days.

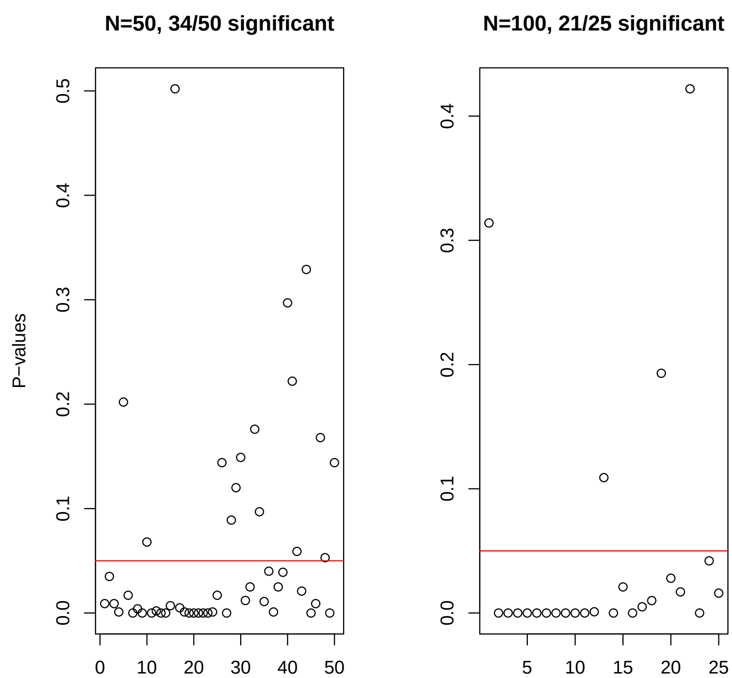

Figure 3.3. P-values for consecutive segments of length N of the price curves $P_n(t)$ of the Disney stock computed using T_N with $v = .9$. The horizontal line shows the 5% threshold.

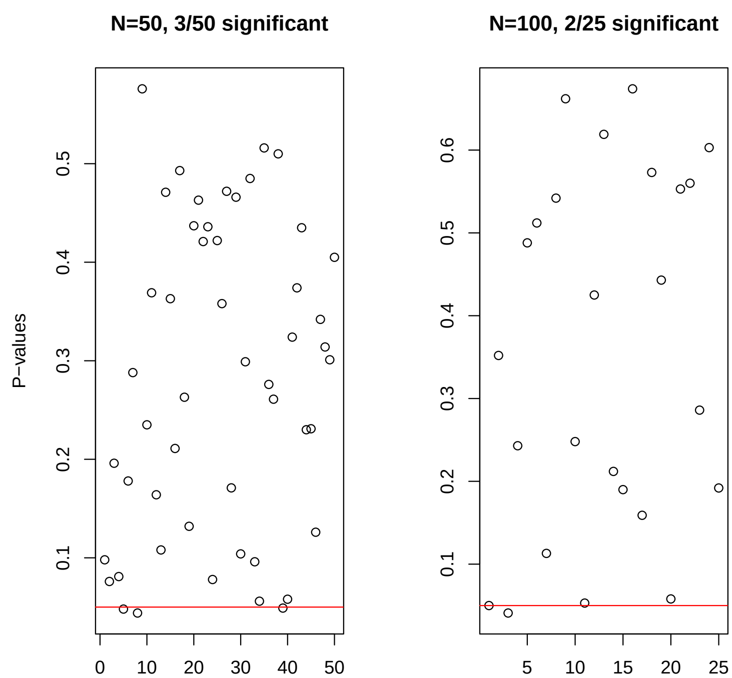

Figure 3.4. P-values for consecutive segments of length N of the CIDR curves $R_n(t)$ for the Disney stock. The red line shows the 5% threshold.

Table 3.1. Empirical sizes for the iid BM and FAR(1) DGPs. We used $h = N^{1/2}$ and the flat–top kernel (3.50). The standard error is approximately 0.9% for the 10% level and 0.4 % for the 5% level.

DGP		BM				FAR(1)			
N		100		250		100		250	
Nominal		10%	5%	10%	5%	10%	5%	10%	5%
Statistics		T_N							
v	.85	13.9	5.6	13.3	5.3	12.3	4.2	11.8	4.1
	.90	12.6	5.4	12.5	5.2	11.8	3.7	11.0	4.1
	.95	12.7	4.6	12.0	4.9	11.3	3.6	10.4	3.7
Statistics		M_N							
v	.85	9.7	2.2	11.4	3.6	11.4	2.3	11.8	4.2
	.90	8.8	1.5	10.5	3.0	9.4	1.6	11.2	4.0
	.95	8.2	0.9	9.9	2.9	8.5	1.2	10.1	3.4
Statistics		T_N^*							
v	.85	11.3	4.9	10.8	4.6	10.2	3.4	10.3	3.7
	.90	11.2	4.4	10.7	4.8	10.0	3.4	10.3	3.4
	.95	11.8	4.4	11.1	4.5	10.6	3.2	10.0	3.6
Statistics		M_N^*							
v	.85	6.3	0.7	8.8	2.6	11.0	2.1	11.2	4.2
	.90	7.0	0.8	8.9	2.3	8.9	1.3	10.7	4.0
	.95	6.9	0.7	8.8	2.7	8.2	1.1	10.0	3.2
Statistics		T_N^0							
v	.85	10.4	3.8	10.3	3.9	10.1	2.8	9.2	3.3
	.90	9.2	2.3	9.0	2.8	7.7	1.5	8.6	2.9
	.95	4.6	0.8	7.6	1.4	5.0	0.1	7.2	1.3
Statistics		M_N^0							
v	.85	6.1	0.5	6.7	2.1	6.7	1.5	7.7	2.8
	.90	4.2	0.8	5.4	1.7	5.8	1.0	7.2	2.4
	.95	2.9	0.3	5.6	1.4	3.3	0.0	5.1	0.9
Statistics		$T_N^0(AM)$							
v	.85	11.9	5.4	10.2	5.1	12.1	7.1	11.7	6.1
	.90	10.3	5.7	9.2	4.8	11.7	7.2	9.8	4.9
	.95	9.9	4.3	9.0	4.7	11.2	6.9	9.7	5.3
Statistics		$M_N^0(AM)$							
v	.85	8.8	5.1	10.7	4.6	12.7	7.7	10.8	5.8
	.90	8.6	5.3	10.0	4.5	12.1	7.3	10.5	5.4
	.95	8.5	4.7	9.8	5.2	11.9	7.1	10.6	5.4

Table 3.2. Empirical power for change point and $I(1)$ alternatives. We used $h = N^{1/2}$ and the flat–top kernel (3.50).

DGP		Change point				I(1)			
N		100		250		100		250	
Nominal		10%	5%	10%	5%	10%	5%	10%	5%
Statistic		T_N							
v	.85	80.7	56.4	99.6	98.1	99.3	96.5	99.2	96.3
	.90	80.1	56.6	99.5	97.6	99.4	95.8	99.2	96.1
	.95	79.2	54.4	99.4	97.6	99.1	96.2	99.2	96.3
Statistic		M_N							
v	.85	50.6	14.7	95.2	84.1	93.8	68.3	97.7	92.5
	.90	46.7	11.0	94.7	82.7	92.8	64.5	97.6	92.4
	.95	79.2	54.4	99.4	97.6	90.9	61.4	99.2	96.3
Statistic		T_N^*							
v	.85	77.2	52.1	99.3	97.6	98.9	95.7	98.0	94.2
	.90	77.8	54.3	99.5	97.5	99.2	95.8	98.4	95.7
	.95	77.5	53.7	99.4	97.6	99.1	96.0	99.1	96.1
Statistic		M_N^*							
v	.85	39.6	8.5	93.7	78.1	93.5	67.9	94.9	88.2
	.90	39.9	7.7	93.9	79.6	92.7	63.9	96.2	89.3
	.95	40.8	6.8	94.5	79.8	90.5	61.2	96.6	90.0
Statistic		T_N^0							
v	.85	85.8	55.1	99.8	98.9	99.5	98.1	98.6	96.2
	.90	86.6	52.0	100	99.6	99.7	98.8	99.3	98.7
	.95	74.7	31.3	99.9	98.4	100	96.0	99.9	99.8
Statistic		M_N^0							
v	.85	35.0	7.8	97.2	77.7	86.1	75.2	97.9	92.7
	.90	31.0	5.9	98.0	71.5	90.9	73.4	99.2	95.4
	.95	21.1	4.8	93.0	63.0	96.8	75.9	100	98.5
Statistic		$T_N^0(AM)$							
v	.85	96.6	91.6	100	100	99.6	99.3	99.8	99.7
	.90	94.9	85.5	100	100	100	99.9	100	100
	.95	82.5	70.0	100	100	100	100	100	100
Statistic		$M_N^0(AM)$							
v	.85	85.3	71.3	99.9	99.8	93.4	85.0	99.8	99.3
	.90	68.7	52.3	99.7	98.8	96.3	90.3	99.9	99.9
	.95	43.7	28.5	97.0	92.0	98.6	96.3	100	100

CHAPTER 4

TESTING EQUALITY OF MEANS WHEN THE OBSERVATIONS ARE FROM FUNCTIONAL TIME SERIES [3]

There are numerous examples of functional data in areas ranging from earth science to finance where the problem of interest is to compare several functional populations. In many instances, the observations are obtained consecutively in time, and thus the classical assumption of independence within each population may not be valid. In this chapter, we derive a new, asymptotically justified method to test the hypothesis that the mean curves of multiple functional populations are the same. The test statistic is constructed from the coefficient vectors obtained by projecting the functional observations into a finite dimensional space. Asymptotics are established when the observations are considered to be from stationary functional time series. Although the limit results hold for projections into arbitrary finite dimensional spaces, we show that higher power is achieved by projecting onto the principle components of empirical covariance operators which diverge under the alternative. Our method is further illustrated by a simulation study as well as an application to electricity demand data.

4.1 Introduction

To this day, a frequently used tool to analyze multiple populations is the one–way analysis of variance (ANOVA) in which the means of k populations are compared. For scalar and vector valued data, this problem has been extensively studied under numerous conditions, and we refer to Anderson (2003) for a review of the subject. The methods developed therein are suitable to address a host of modern statistical questions; however, it is of increasing interest to consider data which take the form of functions or curves for which finite dimensional approaches are not appropriate. For this reason, the theory of functional data analysis has been steadily growing in recent years and much effort has been put forth to adapt classical statistical procedures, such as ANOVA, for functional data. In order to

[3] The content of this chapter is based on joint research with Lajos Horváth.

formally state the one–way functional analysis of variance (FANOVA) problem, we assume that we have observations from k functional populations which satisfy the one–way layout design

$$X_{i,j}(t) = \mu_i(t) + \eta_{i,j}(t),\ t \in [0,1],\ 1 \le i \le k \text{ and } 1 \le j \le N_i, \tag{4.1}$$

where $X_{i,j}$ is the j^{th} observation from the i^{th} population, μ_i is the common mean function of the i^{th} population, and $\eta_{i,j}$ is a random error function satisfying

$$E\eta_{i,j}(t) = 0,\ t \in [0,1],\ 1 \le i \le k, \text{ and } 1 \le j \le N_i. \tag{4.2}$$

The assumption that $t \in [0,1]$ is made without loss of generality. We also assume that observations from separate populations are independent, namely

$$\text{the error sequences } \ \{\eta_{i,j}, 1 \le j \le N_i\} \ \text{ are independent.} \tag{4.3}$$

We wish to test the null hypothesis $H_0 : \mu_1(\cdot) = \mu_2(\cdot) = \ldots = \mu_k(\cdot)$, where equality holds in the L^2 sense, versus the general alternative $H_A : H_0$ does not hold. Assuming two independent populations based on independent random functions, Fan and Lin (1998) and Hall and Keilegom (2007) developed testing procedures. Testing for differences between the means of several populations, Laukaitis and Račkauskas (2005), Abramovich and Angelini (2006), Antoniadis and Sapatinas (2007), and Martínez–Camblor and Corral (2011) expand the observations using wavelets and the tests are based on the wavelet coefficients. Due to the complexity of the distribution of the test statistics used, the critical values are obtained by resampling. Cuevas et al. (2004) developed a test statistic using the L^2 norm of the difference between the population means and the total mean, which was a direct analog of the classical F–test. Again in this case, the critical values of the test statistic could not be computed explicitly, and a Monte–Carlo method is proposed. Many of these results are summarized in the recent book of Zhang (2013). For some applications of functional analysis of variance, we refer to Hooker (2007) and Drignei (2010).

Outside of the difficulty to implement resampling and Monte–Carlo procedures involved in many of the currently available tests, the assumption of independence within each population is often not valid for functional data which are obtained sequentially over time, i.e. as observations of a functional time series. This is frequently the case for econometric and geological data where functional data objects are often obtained by dividing long, continuous records into smaller hourly or daily observations. The problem of testing the homogeneity of mean curves from dependent functional populations has received little attention other than the two sample case (cf. Horváth, Kokoszka, and Reeder (2013)). The goal of the present chapter is to provide a new asymptotic test of H_0 which is both easy to implement, by not relying on resampling or Monte–Carlo methods to obtain critical values, and also accommodates the case of weak dependence within each functional population.

The rest of the chapter is organized as follows. In Section 4.2, we introduce the mathematical framework used to model the errors terms in (4.1) and develop test statistics along with their Gaussian asymptotics under H_0. In Section 4.3, we discuss the behavior of the test statistics under the alternative. Section 4.4 contains a simulation study which explains the implementation of the testing procedure and investigates the size and power in case of finite sample sizes. Section 4.5 provides a detailed study of the electricity demand in Adelaide, Australia. The proofs of the main results are given in Sections 4.6 and 4.7. Some technical lemmas used in Sections 4.6 and 4.7 are in the appendices.

4.2 Main results

Throughout this chapter, the maximum norm of vectors and matrices will be denoted by $|\cdot|$ and the L^2 norm is given by $\|h\|= (\langle h, h\rangle)^{1/2}$, where $\langle f, h\rangle = \int f(t)h(t)dt$ denotes the inner product. We write $\int$ for $\int_0^1$. In order to lighten the notation, we write f for $f(t)$ when it does not cause confusion. To establish the asymptotic properties of the test statistics proposed below, we will require that the error terms $\eta_{i,j}$ of (4.1) be in the class of L^2 m–decomposable Bernoulli shifts which are defined as follows: Let $\boldsymbol{\eta}_i = \{\eta_{i,j}(t)\}_{j=-\infty}^{\infty}, 1 \leq i \leq k$. We assume that for all $1 \leq i \leq k$

$\boldsymbol{\eta}_i$ forms a sequence of Bernoulli shifts, i.e. $\eta_{i,j}(t) = g_i(\epsilon_{i,j}, \epsilon_{i,j-1}, ...)(t)$ (4.1)
for some nonrandom measurable function $g_i : S^\infty \mapsto L^2$ and i.i.d. random innovations $\epsilon_{i,j},\ -\infty < j < \infty$, with values in a measurable space S,

$$\eta_{i,j}(t) = \eta_{i,j}(t, \omega) \text{ is jointly measurable in } (t, \omega)\ (-\infty < j < \infty), \tag{4.2}$$

$$E\|\eta_{i,0}\|^2 < \infty, \tag{4.3}$$

the sequence $\{\boldsymbol{\eta}_i\}$ can be approximated by ℓ–dependent sequences (4.4)
$\{\eta_{i,j,\ell}\}, -\infty < j < \infty, 1 \leq \ell < \infty$ in the sense that

$$\sum_{\ell=1}^{\infty}(E\|\eta_{i,j} - \eta_{i,j,\ell}\|^2)^{1/2} < \infty$$

where $\eta_{i,j,\ell}$ is defined by $\eta_{i,j,\ell} = g_i(\epsilon_{i,j}, \epsilon_{i,j-1}, ..., \epsilon_{i,j-\ell+1}, \boldsymbol{\epsilon}^*_{i,j,\ell})$,
$\boldsymbol{\epsilon}^*_{i,j,\ell} = (\epsilon^*_{i,j,\ell,j-\ell}, \epsilon^*_{i,j,\ell,j-\ell-1}, \ldots)$, where the $\epsilon^*_{i,j,\ell,k}$'s are independent copies of

$\epsilon_{i,0}$, independent of $\{\epsilon_{i,j}, 1 \leq i \leq k, -\infty < j < \infty\}$.

It follows from (4.1) that $\{X_{i,j}\}_{j=-\infty}^{\infty}$ is a stationary and ergodic sequence. The framework outlined for the error terms in (4.1)–(4.4) is very flexible as long as the process is thought to be driven by a sequence of underlying independent innovations. For example, we may take the space S to be $\mathbb{R}$ and $\eta_{i,j}(t) = \sum_{k=0}^{\infty} \epsilon_{i,j-k} f_{i,k}(t)$, where the $f_{i,k}$ are deterministic functions in L^2 decaying in k sufficiently fast to make (4.4) hold. Also, we may take $\eta_{i,j}$ to be a standard linear process in L^2 (cf. Bosq 2000) in which case it would be natural to take $S = L^2$.

Processes satisfying (4.1)–(4.4) are called L^2 m–decomposable processes by Hörmann and Kokoszka (2010). Aue et al. (2012) show that many stationary time series models based on independent innovations satisfy assumptions (4.1)–(4.4). Their examples include autoregressive, moving average and linear processes in Hilbert spaces, the nonlinear functional ARCH(1) model (cf. Hörmann et al. (2012)), and bilinear models (cf. Hörmann and Kokoszka (2010)). Scalar processes satisfying (4.4) were studied in Wu and Min (2005) and the condition serves as an alternative to the classical mixingale assumption; see Hannan (1973). One weakness of the model (4.1)–(4.4) is that it is in practice impossible to know whether dependent errors are driven by underlying independent innovations. Since one of the purposes of assuming (4.1)–(4.4) is to guarantee the central limit theorem for the errors, one could conceive of trading out these assumptions for other weak dependence models, like near epoch dependence or stationary β-mixing (cf. Pötscher and Prucha (1991) and Doukhan et al. (1995), respectively), which allow for the central limit theorem without relying on such structure. For our results, it is required that certain Bartlett type covariance estimators be consistent in L^2, which as of yet has only been established under (4.1)–(4.4). The main idea behind our testing procedure is a dimension reduction technique in which the infinite dimensional observations $X_{i,j}(t)$ are projected onto a suitably chosen finite dimensional space. Suppose this space is spanned by d orthonormal functions $\{\varphi_1, \varphi_2, \ldots, \varphi_d\}$. These functions should be chosen in such a way that both the finite dimensional projections of the $X_{i,j}$s accurately represent the original observations, and the test statistics based on these projections have high power to detect H_A. The former requirement is a familiar one from finite dimensional data analysis and it is typically addressed using principle component analysis. A large amount of research has been done to extend the ideas of principle component analysis to functional data; see Ramsay and Silvermann (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012) for reviews of such work. Following Horváth, Kokoszka and Reeder (2013), instead of working with covariance functions $\mathrm{Cov}(X_{i,1}(t), X_{i,1}(s))$ to generate the principle component basis, our method uses the long-run covariances

$$D_i(t,s) = \mathrm{Cov}(X_{i,1}(t), X_{i,1}(s)) + \sum_{\ell=2}^{\infty}[\mathrm{Cov}(X_{i,1}(t), X_{i,\ell}(s)) + \mathrm{Cov}(X_{i,1}(s), X_{i,\ell}(t))],$$

$1 \le i \le k$. This choice is motivated by the fact that our test statistics are developed from the projections of $\sum_{j=1}^{N_i} X_{i,j}(t)$ and so, since D_i is the asymptotic covariance function of this sum, projecting onto its eigenfunctions gives asymptotically optimal finite dimensional representations. We do not wish to assume, however, that the long-run covariances are homogenous across the populations, and thus we use a weighted sum of the D_is. Suppose that

$$\lim_{N\to\infty} \frac{N_i}{N} = a_i > 0, \text{ where } N = N_1 + N_2 + \ldots + N_k, \tag{4.5}$$

which means that the proportion of observations sampled from each population does not degenerate as the sampling progresses. It follows that the function

$$D(t,s) = \sum_{i=1}^{k} a_i D_i(t,s) \tag{4.6}$$

is non–negative definite so there are eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots \ge 0$ and corresponding orthonormal functions $\{\varphi_\ell\}_{\ell=1}^{\infty}$ satisfying

$$\lambda_i \varphi_i(t) = \int D(t,s)\varphi_i(s)ds \tag{4.7}$$

(cf. Debnath and Mikusiński (2005), p. 186). In order to have uniquely defined eigenvalues and one-dimensional eigenspaces, we assume

$$\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1} > 0, \tag{4.8}$$

which is a widely used assumption in functional data analysis. We would then like to take $\varphi_1, \varphi_2, \ldots, \varphi_d$ of (4.7) as the basis for the projections. Since these functions are defined by the unknown $D(t,s)$, they must be estimated from the sample.

In this chapter, we propose two Bartlett–type estimators for $D_i(t,s)$. Using the combined sample mean $\bar{X}_{\cdot\cdot}(t) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i} X_{i,j}(t)$, we define the sample autocovariance functions

$$\tilde{\gamma}_{i,\ell}(t,s) = \frac{1}{N_i}\sum_{j=\ell+1}^{N_i} \left(X_{i,j}(t) - \bar{X}_{\cdot\cdot}(t)\right)\left(X_{i,j-\ell}(s) - \bar{X}_{\cdot\cdot}(s)\right).$$

and the corresponding long-run covariance function estimators

$$\tilde{D}_{N_i,i}(t,s) = \tilde{\gamma}_{i,0}(t,s) + \sum_{\ell=1}^{N_i-1} K(\ell/h)\left(\tilde{\gamma}_{i,\ell}(t,s) + \tilde{\gamma}_{i,\ell}(s,t)\right).$$

This gives the first estimator of D,

$$\tilde{D}_N(t,s) = \sum_{i=1}^{k} \frac{N_i}{N} \tilde{D}_{N_i,i}(t,s). \tag{4.9}$$

Similarly, using

$$\bar{X}_{i\cdot}(t) = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{i,j}(t),$$

the sample means computed within each population, we define

$$\hat{\gamma}_{i,\ell}(t,s) = \frac{1}{N_i} \sum_{j=\ell+1}^{N_i} \left(X_{i,j}(t) - \bar{X}_{i\cdot}(t)\right)\left(X_{i,j-\ell}(s) - \bar{X}_{i\cdot}(s)\right),$$

$$\hat{D}_{N_i,i}(t,s) = \hat{\gamma}_{i,0}(t,s) + \sum_{\ell=1}^{N_i-1} K(\ell/h)\left(\hat{\gamma}_{i,\ell}(t,s) + \hat{\gamma}_{i,\ell}(s,t)\right)$$

and

$$\hat{D}_{N,p}(t,s) = \sum_{i=1}^{k} \frac{N_i}{N} \hat{D}_{N_i,i}(t,s). \tag{4.10}$$

In an analogy with the simple one–way ANOVA test for univariate data, the estimator $\hat{D}_{N,p}$ is a measure of the within-population covariation and will converge to D regardless of the truth or falsity of H_0. On the other hand, $\tilde{D}_N$ is a measure of overall covariation and will converge to D only under H_0. The kernel K satisfies

$$K(0) = 1, \tag{4.11}$$

$$K(u) = 0 \ \text{ if } \ u > c \ \text{ with some } c > 0, \tag{4.12}$$

and

$$K \ \text{ is continuous on } \ [0,c], \text{ where } c \text{ is given in (4.12)}. \tag{4.13}$$

We require that the window (or smoothing parameter) h satisfy

$$h = h(N) \to \infty \quad \text{and} \quad \frac{h(N)}{N} \to 0, \quad \text{as} \quad N \to \infty. \tag{4.14}$$

To prove the consistency of Bartlett type kernel estimators, the condition

$$\lim_{\ell\to\infty} \ell(E\|\eta_{i,j} - \eta_{i,j;\ell}\|^2)^{1/2} = 0 \tag{4.15}$$

must be assumed in addition to (4.4). Now to form the empirical score vectors, we use the eigenfunctions associated with the first d largest eigenvalues of $\hat{D}_{N,p}$ or $\tilde{D}_N$ defined as

$$\tilde{\lambda}_i \tilde{\varphi}_i(t) = \int \tilde{D}_N(t,s)\tilde{\varphi}_i(s)ds \tag{4.16}$$

and

$$\hat{\lambda}_i \hat{\varphi}_i(t) = \int \hat{D}_{N,p}(t,s)\hat{\varphi}_i(s)ds. \tag{4.17}$$

Let

$$\tilde{\boldsymbol{\xi}}_{i,j} = (\langle X_{i,j}, \tilde{\varphi}_1\rangle, \langle X_{i,j}, \tilde{\varphi}_2\rangle, \ldots, \langle X_{i,j}, \tilde{\varphi}_d\rangle)^T$$

denote the vector of scores which define the projection of $X_{i,j}$ into the space spanned by $\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_d$. We define similarly

$$\hat{\boldsymbol{\xi}}_{i,j} = (\langle X_{i,j}, \hat{\varphi}_1\rangle, \langle X_{i,j}, \hat{\varphi}_2\rangle, \ldots, \langle X_{i,j}, \hat{\varphi}_d\rangle)^T.$$

The averages of the empirical score vectors within each population are defined as

$$\tilde{\boldsymbol{\xi}}_{i\cdot} = \frac{1}{N_i}\sum_{j=1}^{N_i} \tilde{\boldsymbol{\xi}}_{i,j} \quad \text{and} \quad \hat{\boldsymbol{\xi}}_{i\cdot} = \frac{1}{N_i}\sum_{j=1}^{N_i} \hat{\boldsymbol{\xi}}_{i,j},$$

$1 \le i \le k$. We may then estimate the common mean vector of the scores under H_0 with either

$$\tilde{\boldsymbol{\xi}}_{\cdot\cdot} = \left(\sum_{i=1}^{k} N_i \tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1} \sum_{i=1}^{k} N_i \tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\boldsymbol{\xi}}_{i\cdot} \quad \text{or} \quad \hat{\boldsymbol{\xi}}_{\cdot\cdot} = \left(\sum_{i=1}^{k} N_i \hat{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1} \sum_{i=1}^{k} N_i \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\xi}}_{i\cdot},$$

where

$$\tilde{\boldsymbol{\Sigma}}_i = \left\{ \iint \tilde{D}_{N_i,i}(t,s)\tilde{\varphi}_\ell(t)\tilde{\varphi}_j(s)dtds,\ 1 \le j,\ell \le d \right\}$$

and

$$\hat{\boldsymbol{\Sigma}}_i = \left\{ \iint \hat{D}_{N_i,i}(t,s)\hat{\varphi}_\ell(t)\hat{\varphi}_j(s)dtds,\ 1 \le j,\ell \le d \right\},$$

$1 \le i \le k$. The definitions of $\tilde{\boldsymbol{\xi}}_{\cdot\cdot}$ and $\hat{\boldsymbol{\xi}}_{\cdot\cdot}$ assume that $\tilde{\boldsymbol{\Sigma}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ are nonsingular, which holds asymptotically with probability tending to one if

$$\text{for each } 1 \le i \le k \text{ the function } D_i(t,s) \text{ is (strictly) positive definite.} \tag{4.18}$$

Our testing procedure may then be based on either

$$\tilde{T}_N = \sum_{i=1}^{k} N_i \left(\tilde{\boldsymbol{\xi}}_{i\cdot} - \tilde{\boldsymbol{\xi}}_{\cdot\cdot}\right)^T \tilde{\boldsymbol{\Sigma}}_i^{-1} \left(\tilde{\boldsymbol{\xi}}_{i\cdot} - \tilde{\boldsymbol{\xi}}_{\cdot\cdot}\right) \quad \text{or} \quad \hat{T}_N = \sum_{i=1}^{k} N_i \left(\hat{\boldsymbol{\xi}}_{i\cdot} - \hat{\boldsymbol{\xi}}_{\cdot\cdot}\right)^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left(\hat{\boldsymbol{\xi}}_{i\cdot} - \hat{\boldsymbol{\xi}}_{\cdot\cdot}\right).$$

The first result of our chapter shows that both $\tilde{T}_N$ and $\hat{T}_N$ satisfy the same limit theorem under the null hypothesis.

Theorem 4.2.1. *If H_0, (4.1)–(4.5), (4.8), and (4.11)–(4.15) are satisfied, then we have*

$$\hat{T}_N \ \stackrel{\mathcal{D}}{\to}\ \chi^2(d(k-1)) \tag{4.19}$$

and

$$\tilde{T}_N \ \stackrel{\mathcal{D}}{\to}\ \chi^2(d(k-1)), \tag{4.20}$$

where $\chi^2(d(k-1))$ stands for a χ^2 random variable with $d(k-1)$ degrees of freedom.

The proof of Theorem 4.2.1 is postponed to Section 4.6.

4.3 Consistency of the test statistics

Let

$$\bar{\mu}(t) = \sum_{i=1}^{k} a_i \mu_i(t).$$

If the projections are defined via the eigenfunctions defined in (4.16), the condition for consistency is simple.

Theorem 4.3.1. *If H_A, (4.1)–(4.5), (4.8), and (4.11)–(4.15) are satisfied and*

$$\langle \mu_i - \bar{\mu}, \varphi_i \rangle \neq 0 \ \text{ for some } \ 1 \leq i \leq d, \tag{4.1}$$

then $\tilde{T}_N \to \infty$ in probability.

No condition like (4.1) is needed, however, if the projection functions are the eigenfunctions of $\tilde{D}_{N,p}$. We show in Section 4.7 that $\tilde{D}_N(t,s)$ is close to

$$D_N^*(t,s) = \left(2\sum_{\ell=1}^{\infty} K(\ell/h)\right) \sum_{i=1}^{k} M_i(t)M_i(s) + D(t,s), \tag{4.2}$$

where

$$M_i(t) = \mu_i(t) - \bar{\mu}(t). \tag{4.3}$$

Since $\sum_{i=1}^{k} M_i(t)M_i(s)$ is a non-negative definite function, one can find orthonormal functions $\kappa_1, \kappa_2, \ldots, \kappa_m, m \leq k-1$ and positive numbers $e_1, e_2, \ldots, e_m$ such that

$$\sum_{i=1}^{k} M_i(t)M_i(s) = \sum_{i=1}^{m} e_i \kappa_i(t)\kappa_i(s).$$

In order that the eigenfunctions of D_N^* be identifiable up to a sign, we assume that

$$e_1 > e_2 > \ldots > e_m. \tag{4.4}$$

Since $\sum_{\ell=1}^{\infty} K(\ell/h) \to \infty$, as $h \to \infty$, the first m largest eigenvalues and eigenfunctions of D_N^* will be determined asymptotically by the term $\sum_{i=1}^{m} e_i \kappa_i(t)\kappa_i(s)$. Let $\mathcal{A}_0$ denote the span of $\kappa_1, \ldots, \kappa_m$ and $\bar{\mathcal{B}}$ be the orthogonal complement of the set $\mathcal{B}$. We say that D has regular maxima of order $d-m$ with respect to $\mathcal{A}_0$ if there exist constants $r_1 > r_2 > \ldots > r_{d-m} > 0$ and orthonormal functions $g_1, g_2, \ldots, g_{d-m}$ such that with $\mathcal{A}_i = \mathrm{span}(\kappa_1, \ldots, \kappa_m, g_1, \ldots, g_i), 1 \le i \le d-m$,

$$r_i = \sup_{g \in \bar{\mathcal{A}}_{i-1}: \|g\|=1} \iint g(t)D(t,s)g(s)dtds = \iint g_i(t)D(t,s)g_i(s)dtds \;\; 1 \le i \le d-m.$$

We note that the functions $g_1, g_2, \ldots, g_{d-m}$ are unique up to signs.

Theorem 4.3.2. *If H_A, (4.1)–(4.5), (4.11)–(4.15), and (4.4) are satisfied, D has regular maxima of order $d-m$ with respect to $\mathcal{A}_0$ and $h/N^{1/2} \to 0$, then $\hat{T}_N \to \infty$ in probability.*

If $d \le m$, i.e. the number of principle components used is less than the dimension of the span of the population means, then there is no restriction on the covariance function D. Notice that $1 \le m \le k-1$.

It is a standard assumption in functional principle component analysis that the eigenfunctions of a covariance operator are associated with unique eigenvalues. Therefore, the assumption that D has regular maxima is along the lines of standard assumptions in the literature. Also, under the assumptions of Theorem 4.3.2, we have that there are no ties among the first d largest eigenvalues of D_N^*.

4.4 Implementation of the test and a simulation study

The biggest obstacle in the implementation of Theorem 4.2.1 is the estimation of the long-run covariance kernel in (4.9) and (4.10). Several issues must be considered. The choice of the kernel $K(\cdot)$ and the smoothing bandwidth h are the most obvious. The issues of bandwidth and kernel selection have been extensively studied in the statistical literature over the last three decades. For scalar data, perhaps the best known contributions are those of Andrews (1991) and Andrews and Monahan (1992) who introduced data-driven bandwidth selection techniques. While these approaches possess optimality properties in general regression models with heteroskedastic and correlated errors, they are not optimal in all specific applications. This chapter focuses on the derivation and large sample theory for a functional analysis of variance test, and hence we cannot present here a comprehensive study of the finite sample properties of the covariance function estimation, which are still being investigated for scalar time series. We do wish to offer some practical guidance on this aspect of the procedure and report approaches which worked well for the data-generating processes we considered. It is argued in Politis (2003) that the flat top kernel

$$K(t) = \begin{cases} 1, & 0 \leq t < 0.1 \\ 1.1 - |t|, & 0.1 \leq t < 1.1 \\ 0, & |t| \geq 1.1. \end{cases} \tag{4.1}$$

has better properties than the Bartlett or the Parzen kernels due to its smaller bias. In our results, we used the flat top kernel of (4.1). Following the arguments in Andrews (1991), one can show that the optimal bandwidth h in terms of minimizing the integrated mean squared error has the form $h^* = c_B N^{1/3}$ for the Bartlett kernel and $h^* = c_P N^{1/5}$ for the Parzen kernel. The constants c_B and c_P not only depend on the kernels but they are very difficult functionals of the higher order covariance structure of $\{X_{i,j}, 1 \leq j < \infty\}$. Our simulations showed that $h = N^{1/4}$ proves satisfactory when the observations are independent or weakly dependent (functional autoregressive processes) and the results (empirical sizes and power functions) are stable for this choice of h, i.e. the results of the simulations change little for small changes in h. Throughout this section, we used $h = N^{1/4}$ whenever the flat top kernel was used. Once the kernel and the bandwidth have been selected, the empirical eigenvalues and eigenfunctions can be computed using (4.16) or (4.17). To select d, we use the standard "cumulative variance" approach recommended by Ramsay and Silverman (2005) and Horváth and Kokoszka (2012); d is computed so that roughly v% of the sample variance is explained by the first d principal components. In the independent case, for example, this amounts to taking $d = d_v$ such that

$$\frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{d_v}}{\int \hat{D}_{N,p}(t,t)dt} \approx v,$$

where the $\hat{\lambda}_i$'s are defined in (4.17) and $\hat{D}_{N,p}$ is defined in (4.9). A general recommendation is to use v equal to .90. This is the choice we have made for the analysis below. The results of the simulations and applications below changed little for other values of v between .80 and .95. All simulations and the calculations for the applications were performed using the R programming language (R development core team (2008)).

4.4.1 Finite sample size

Using a simulation study, we first compare the empirical size of the test implemented as described above. Under H_0, we can assume without loss of generality that in (4.1) $\mu_i(t) = 0$, and hence we use $X_{i,j} = \eta_{i,j}$. We consider three data-generating processes (DGPs) to generate the errors terms $\eta_{i,j}$. In the case IID we use the sequence

$$\eta_{i,j}(t) = B_{i,j}(t), \;\; t \in [0,1], \;\; 1 \leq i \leq k, \;\; 1 \leq j \leq N, \tag{4.2}$$

where $\{B_{i,j}\}_{j=1}^N$, $1 \leq i \leq k$ are independent identically distributed Brownian bridges. To study the size of the test when the populations exhibit temporal dependence, we consider error sequences which follow the functional autoregressive process of order one defined by the equation

$$\eta_{i,j}(t) = \int f(t,u)\eta_{i,j-1}(u)du + B_{i,j}(t), \quad t \in [0,1], \; 1 \leq i \leq k, \; 1 \leq j \leq N, \tag{4.3}$$

which we refer to as $\text{FAR}_f = \text{FAR}_f(1)$. In order to generate $\eta_{i,0}$, we use a burn in sample of length 100 according to (4.3) which starts from an independent innovation. It is shown in Bosq (2000) that if $||f||< 1$, then (4.3) has a unique stationary and ergodic solution. In this section, we will consider two different kernels: $\psi_1(t,s) = \min(t,s)$ and $\psi_2(t,s) = c\exp\left((t^2+s^2)/2\right)$. The simulations are performed with $c = .3416$ so that $||\psi_2||\approx 1/2$. For comparison $||\psi_1||\approx 1/3$. To obtain the results below, we used the empirical projection functions defined in (4.17), respectively. Each DGP was simulated one thousand times independently for each value of N, and the percentage of rejections of the null hypothesis is reported when the significance levels are 10%, 5% and 1% in Table 4.1. Based on these results, we reach the following conclusions:

(1) When the populations are comprised of independent observations, the test exhibits close to nominal sizes even for small values of N. Also the value of k has little effect on the size in this case.

(2) When the samples are comprised of dependent observations, the size of the test may be inflated for small values of N or when k is large and/or the dependence within the populations is strong. This problem is ameliorated by taking a larger sample and good size can be achieved even for large k and fairly strong dependence.

(3) If it can be assumed that the population samples under study are comprised of independent or weakly dependent observations, then we have no reservations about the use of the test. If the temporal dependence within the samples is not too strong and k is small, then the test may be applied to the data without reservations. However, if the dependence

within the samples is suspected to be quite strong and/or the number of populations under study is large, then we recommend that the test only be applied if a sufficiently large sample may be obtained.

We also investigate the size of the test in finite samples when the underlying population covariance functions are heterogeneous. A large number of possible DGPs could be considered here. We focus on two examples, one for each of the independent and dependent cases, which illustrate the theory presented above and provide new insights relative to the simulations we have already performed. In our own studies in the case of heterogeneous covariances, we observed similar behavior as above when the number of populations k was changed, and thus we hold $k = 3$ for this case. In the heterogenous independent case (HIID), we will assume that $k = 3$, $\eta_{1,j}(t) = W_j(t)$, $\eta_{2,j}(t) = B_{1,j}(t)$, and $\eta_{3,j}(t) = 2B_{2,j}(t)$, where $t \in [0, 1]$, $1 \leq j \leq N$, $\{W_j\}_{j=1}^N$ are i.i.d. Wiener processes, and $B_{i,j}$ are defined above. In the heterogenous dependent case (HDEP), we take the sequences $\eta_{1,j}$ following a FAR_{ψ_1}, $\eta_{2,j}$ following a FAR_{ψ_2}, and $\eta_{3,j}(t) = .5\eta_{3,j-1}(t) + B_{3,j}(t)$, where as before a burn in sample of 100 was used starting from an independent innovation. In each case, the covariances differ significantly across the three populations. Again each DGP was simulated 1000 times and the percentage of rejections of H_0 is reported when the nominal levels were 10%, 5% and 1%. The results are given in Table 4.2, from which we can draw the following conclusions:
(1) If the observations form a simple random sample, then heterogeneous covariances may inflate the size of the test slightly; however, the empirical rejection rate was close to the nominal levels even for small values of N. Similar conclusions were drawn for other values of k which we considered ($4 \leq k \leq 10$).
(2) If the observations exhibit temporal dependence, then good size is achieved in the case of heterogeneous covariances as long as the sample size is sufficiently large and the dependence within each population is not too strong. If either of these two conditions are not met, then the size may be inflated.

4.4.2 Power study

We now turn to the study of the power of the test in finite samples. The number of possible alternatives which could be considered to study the power of the FANOVA test is enormous, since the variables for the experiment include the choices for the population mean functions under H_A, how much dependence is allowed within the populations, and how many populations k which are used. Given the consideration of space, we cannot possibly pursue even a fraction of these possible alternatives here, and thus we focus in this section on a few examples which seem plausible for real data and highlight the most interesting results of our own more thorough simulations. It was confirmed by our simulations that in most situations, the power of the test is increasing with k. In this section, we focus on the "worst case" other than the well studied two sample problem of $k = 3$. We consider five different regimes for the mean curves themselves which satisfy H_A: $(M1)$ $\mu_i(t) = t(1-t)$ if $i = 1$ and 0 if $i = 2, 3$; $(M2)$ $\mu_i(t) = .1\sin(i\pi t),\ i = 1, 2, 3$; $(M3)$ $\mu_i(t)$ is $t^5(1-t)$, if $i = 1$, $t^3(1-t)^3$, if $i = 2, t(1-t)^5$, if $i = 3$; $(M4)$ $\mu_i(t) = i/10$, if $i = 1, 2, 3$; $(M5)$ $\mu_i(t) = i/20,\ i = 1, 2, 3$ for $t \in [0, 1]$. Case $(M1)$ corresponds to the situation where only one of the mean curves is different from the others. In Case $(M2)$, the mean curves each fluctuate around zero by the same scale but exhibit different frequencies. In Case (M3), the mean curves exhibit small deviations, however, they have vastly different shapes. In Cases $(M4)$ and $(M5)$, the mean curves are simply different scalars, yet these represent the realistic data scenario when the differences among observed curves across some functional populations are simply level shifts. The observations are then constructed using expression (4.1) with $\eta_{i,j}(t)$ following both (4.2) and (4.3). In the interest of space, under (4.3), we only consider the kernel ψ_2, since this is the example with more dependence and gives more contrast with the i.i.d. example. We used the empirical projection functions defined in (4.16). Again each data-generating process is simulated 1000 times and the rejection rate of H_0 is reported when the significance level is 10%, 5%, and 1%. Table 4.3 shows the results of these simulations. We summarize the results as follows. When the errors are independent, the test has good power to detect even small deviations in the mean curves. When the magnitude of the deviations between the mean curves were as small as 1/20, which is around 1/5th the size of the median of $\sup_{0\le t\le 1}|B(t)|$, the test exhibited good power to detect H_A given a large enough sample size. As expected, the power of the test was not as high in case of dependent errors. However, when the magnitude of the deviations between the mean curves is moderate to large, the test still performs well despite fairly strong dependence. If the magnitude of the deviations between the means is small and the sequence exhibits strong dependence, then a large sample is necessary to detect the difference. The power is improved when the dependence in the error sequence is weakened.

4.5 Applications–electricity demand in Adelaide, Australia

In order to illuminate how our test may be used, we consider a real data example of the daily electricity demand curves constructed from half–hourly measurements of the electricity demand in Adelaide, Australia from 7/6/1997 to 3/31/2007. These data have been made publicly available by Shang and Hyndman (2013) as part of the fds package in R. Although each day is only comprised of 48 observations, it is easy to imagine that these observations are simply a sample from an underlying continuous curve which represents the electricity demand throughout the day, and hence it is more appropriate to perform functional data analysis on the continuous curves constructed by interpolating the data points than using multivariate analysis. Multivariate analysis in this context ignores the fact that the measurements are taken from an underlying curve. Five of such curves constructed using linear interpolation are shown in Figure 4.1. It is argued by Magnano and Boland (2007) that the cost of unserved energy can be valued at thousands of dollars per MWh, and hence there is an incentive to develop accurate models of the daily demand in order to reduce excess electricity generation. A cursory examination of the data set confirms that there are many patterns to be found in the daily demand curves. For example, on the weekdays (Monday through Friday) there seems to be a fairly stable pattern where the lowest demand is observed around 3:30am to 4:30am and the peak demand occurs around 3:30pm. In contrast, on the weekends, the lowest demand seems to come a bit later at around 5:00am on average and the peak demand is not quite as high as on the weekdays and typically occurs later at around 4:30pm. Furthermore, there seem to be differing trends in the demand according to the season. The functional analysis of variance test can be used to shed some light on these data by providing a significance test to differentiate which days and which seasons exhibit differing electricity demand curves on the average. Instead of working with the demand curves themselves, we work with the log differenced demand curves defined below.

Definition 4.5.1. *Suppose $D_n(t)$ is the electricity demand at time t on day n for $t \in [0,1]$, $n = 1, \ldots, N$. The functions $R_n(t) = \ln D_n(t) - \ln D_n(0)$, $t \in [0,1]$, $n = 1, \ldots, N$, are called the* log differenced demand curves *(LDDCs).*

The LDDCs computed from the five curves in Figure 4.1 are shown in Figure 4.2. Since the LDDCs have the same shape as the demand curves, they make suitable substitutes for examining the patterns in the daily fluctuations of electricity demand. The reasons for working with the LDDCs are as follows. Due to an overall linearly increasing trend in the electricity demand over the observation period and the effects of seasonal and acute changes in temperature, the demand curves themselves do not appear to be stationary. In the LDDCs, level stationarity is enforced since each curve starts from zero. Furthermore, taking the logarithm helps to control any scale inflation.

To begin, we consider the problem of testing if the mean of the LDDCs is homogeneous across the four predominant seasons in Adelaide: Summer (December, January, February), Fall (March, April, May), Winter (June, July, August), and Spring (September, November, December). Towards this end, we divided the data set consisting of 3556 daily curves into these four seasonal groups depending on the month in which the observation was taken. From this sample, the observations corresponding to the weekends were removed since the demand behavior is vastly different on these days. After removing the weekends, in total, there are 642 observations from the Spring months, 628 from Fall, 630 from Summer, and 640 from Winter. The mean functions from these samples are shown in Figure 4.3. To allow for dependence within the samples, we implement the test outlined in Section 4.3. When the FANOVA test is applied to these four populations, the test rejects the null hypothesis with a p–value which is less than 10^{-6}. By examining Figure 4.3, it appears that Spring and Fall have similar mean LDDCs. The approximate p–value of the test when applied to just the Spring and Fall samples is approximately .21, indicating that there is not sufficient evidence present in the data to reject the notion that Spring and Fall have the same demand patterns. These findings are consistent with the prevailing theory that the electricity demand is driven mainly by the temperature (see Magnano et al. (2008)). By implementing a Bonferroni type procedure, the global error rate of sequential tests of this nature could be controlled; we however do not pursue such an implementation here.

In order to study whether the daily pattern in electricity demand is homogeneous across each day of the week we divided the data set into seven groups each of size 508 corresponding to the days of the week, Sunday through Saturday, and then computed their LDDCs. Due to the prior analysis of the seasonal trend above, we further grouped the data into the four seasonal groups of Summer, Fall, Winter, and Spring; each subsample for each day contained at least 120 curves. The results of the dependent version of the FANOVA test applied to these samples are displayed in Table 4.4.

We summarize the results as follows. Since in every sample which included Saturday and Sunday, not withstanding the sample which contained just these days, the homogeneity of the mean curves was rejected, we conclude that there is strong statistical evidence that Saturday and Sunday have different demand patterns than any other day of the week. When the test was applied to the samples including the middle weekdays Tuesday, Wednesday, and Thursday, we did not reject the homogeneity of the mean curves, and hence the data suggests that these days have the same demand patterns. When the test was applied to the samples containing each of the weekdays, we emphatically reject the homogeneity of the mean curves given the p–values which are very close to zero in each season. Further insight is gained by examining the results of the test when applied to the samples of the weekdays excluding Monday and Friday. When Friday is excluded, we see that in three out of the four seasons, the test cannot distinguish Monday's mean curve from those of Tuesday, Wednesday, and Thursday at the 5% level. In contrast, when Monday is excluded, the test rejects the homogeneity of the mean curves at the 5% level in three out of four samples and in the fourth sample, the p–value is approaching significance. In summary, the data suggest that the mean electricity patterns on the first four weekdays Monday through Thursday are indistinguishable, and that the patterns on Friday, Saturday, and Sunday are significantly different. These results are expected of course, as social behavior is likely to change on the weekend.

4.6 Proof of Theorem 4.2.1

This proof utilizes central limit theory for dependent sequences of random variables that are Bernoulli shifts. The proof can be adapted to other dependent sequences for which the central limit theorem holds. Let $\mu(\cdot)$ denote the common mean under H_0. Using the central limit theorems for dependent functions in Berkes et al. (2013) and Jirak (2013), we can define k independent Gaussian processes $\Gamma_{1,N_1}(t), \Gamma_{2,N_2}(t), \ldots, \Gamma_{k,N_k}(t)$ with covariance functions D_i such that

$$\max_{1\le i\le k}\left\|\frac{1}{N_i^{1/2}}\sum_{j=1}^{N_i}[X_{i,j}(t)-\mu(t)]-\Gamma_{i,N_i}(t)\right\| = o_P(1). \tag{4.1}$$

Define $\boldsymbol{\mu} = (\langle \mu, \varphi_1\rangle, \langle \mu, \varphi_2\rangle, \ldots, \langle \mu, \varphi_d\rangle)^T$, $\tilde{\boldsymbol{\mu}} = (\langle \mu, \tilde{\varphi}_1\rangle,\ \langle \mu, \tilde{\varphi}_2\rangle, \ldots, \langle \mu, \tilde{\varphi}_d\rangle)^T$, and $\hat{\boldsymbol{\mu}} = (\langle \mu, \hat{\varphi}_1\rangle, \langle \mu, \hat{\varphi}_2\rangle, \ldots, \langle \mu, \hat{\varphi}_d\rangle)^T$ to be the d dimensional projections of $\mu(\cdot)$ onto the theoretical and empirical eigenfunctions, respectively. It follows from Horváth, Kokoszka, and Reeder (2013) that under conditions (4.1)–(4.4) and (4.11)–(4.15)

$$\|\hat{D}_{N,p} - D\| = o_P(1) \quad \text{and} \quad \|\tilde{D}_N - D\| = o_P(1). \tag{4.2}$$

Hence if the λ_i's satsfy (4.8), then we have immediately

$$\max_{1\le i\le d}\|\tilde{\varphi}_i - \hat{c}_i\varphi_i\| \ \xrightarrow{P}\ 0 \ \text{ with } \ \tilde{c}_i = \text{sign}(\langle \tilde{\varphi}_i, \varphi_i\rangle) \tag{4.3}$$

and similarly

$$\max_{1\le i\le d}\|\hat{\varphi}_i - \hat{c}_i\varphi_i\| \ \xrightarrow{P}\ 0 \ \text{ with } \ \hat{c}_i = \text{sign}(\langle \hat{\varphi}_i, \varphi_i\rangle). \tag{4.4}$$

Let $\hat{\mathbf{C}} = \text{diag}(\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_d)$, $\boldsymbol{\xi}_{i,j} = (\langle X_{i,j}, \varphi_1\rangle, \langle X_{i,j}, \varphi_2\rangle, \ldots, \langle X_{i,j}, \varphi_d\rangle)$ denote the projections of the observations and $\boldsymbol{\xi}_{i\cdot} = N_i^{-1}\sum_{1\le j\le N_i} X_{i,j}$. The approximation in (4.1) with (4.4) yields that under H_0

$$\left| N_i^{1/2} \left(\hat{\boldsymbol{\xi}}_{i\cdot} - \bar{\boldsymbol{\mu}} - \widehat{\mathbf{C}}(\boldsymbol{\xi}_{i\cdot} - \boldsymbol{\mu}) \right) \right| \leq \max_{1 \leq \ell \leq d} N_i^{-1/2} \left| \langle \sum_{j=1}^{N_i} (X_{i,j} - \mu), \hat{\varphi}_\ell - \hat{c}_\ell \varphi_\ell \rangle \right| \tag{4.5}$$
$$\leq \max_{1 \leq \ell \leq d} \left\| N_i^{-1/2} \sum_{j=1}^{N_i} (X_{i,j} - \mu) \right\| \|\hat{\varphi}_\ell - \hat{c}_\ell \varphi_\ell\|$$
$$= o_P(1).$$

It also follows from (4.1) that under H_0, we have that

$$\left| N_i^{1/2} (\boldsymbol{\xi}_{i\cdot} - \boldsymbol{\mu}) - \mathfrak{S}_{i,N_i} \right| \leq \max_{1 \leq \ell \leq d} \left| \langle N_i^{-1/2} \sum_{j=1}^{N_i} [X_{i,j} - \mu] - \Gamma_{i,N_i}, \varphi_\ell \rangle \right| \tag{4.6}$$
$$\leq \max_{1 \leq \ell \leq d} \left\| N_i^{-1/2} \sum_{j=1}^{N_i} [X_{i,j} - \mu] - \Gamma_{i,N_i} \right\| \|\varphi_\ell\|$$
$$= o_P(1),$$

where $\mathfrak{S}_{i,N_i} = (\mathcal{S}_{i,N_i}(1), \mathcal{S}_{i,N_i}(2), \ldots, \mathcal{S}_{i,N_i}(d))^T$ with $\mathcal{S}_{i,N_i}(\ell) = \langle \Gamma_{i,N_i}, \varphi_\ell \rangle, 1 \leq i \leq k, 1 \leq \ell \leq d$. Putting together (4.5) and (4.6), we conclude

$$\left| N_i^{1/2} (\hat{\boldsymbol{\xi}}_{i\cdot} - \bar{\boldsymbol{\mu}}) - \widehat{\mathbf{C}} \mathfrak{S}_{i,N_i} \right| = o_P(1). \tag{4.7}$$

It is clear that the vectors $\mathfrak{S}_{i,N_i}, 1 \leq i \leq k$ are independent normal random vectors in $\mathbb{R}^d$ with $E\mathfrak{S}_{i,N_i} = \mathbf{0}$ and $E\mathfrak{S}_{i,N_i}\mathfrak{S}_{i,N_i}^T = \boldsymbol{\Sigma}_i, 1 \leq i \leq k$, where

$$\boldsymbol{\Sigma}_i = \left\{ \iint D_i(t,s) \varphi_\ell(t) \varphi_j(s) dt ds, \ 1 \leq j, \ell \leq d \right\} \ \ 1 \leq i \leq d.$$

Assumption (4.18) and Lemma 4.9.1 imply that $\boldsymbol{\Sigma}_i$ is nonsingular. It follows from (4.4) and the ergodic theorem that for $1 \leq \ell, m \leq d$,

$$\left| \frac{1}{N_i} \sum_{j=1}^{N_i} \langle X_{i,j}, \hat{\varphi}_\ell \rangle \langle X_{i,j}, \hat{\varphi}_m \rangle - \hat{c}_\ell \hat{c}_m \frac{1}{N_i} \sum_{j=1}^{N_i} \langle X_{i,j}, \varphi_\ell \rangle \langle X_{i,j}, \varphi_m \rangle \right|$$
$$\leq \frac{1}{N_i} \sum_{j=1}^{N_i} |\langle X_{i,j}, \hat{\varphi}_\ell \rangle \langle X_{i,j}, \hat{\varphi}_m - \hat{c}_m \varphi_m \rangle| + \frac{1}{N_i} \sum_{j=1}^{N_i} |\langle X_{i,j}, \hat{\varphi}_\ell - \hat{c}_\ell \varphi_\ell \rangle \langle X_{i,j}, \hat{c}_m \varphi_m \rangle|$$
$$\leq \|\hat{\varphi}_m - \hat{c}_m \varphi_m\| \frac{1}{N_i} \sum_{j=1}^{N_i} \|X_{i,j}\|^2 + \|\hat{\varphi}_\ell - \hat{c}_\ell \varphi_\ell\| \frac{1}{N_i} \sum_{j=1}^{N_i} \|X_{i,j}\|^2$$
$$= o_P(1),$$

and therefore,

$$\left| \hat{\boldsymbol{\Sigma}}_i - \hat{\mathbf{C}}_i \boldsymbol{\Sigma}_i \hat{\mathbf{C}}_i \right| = o_P(1). \tag{4.8}$$

Using (4.7) and (4.8), we get that

$$\left|(\hat{\boldsymbol{\xi}}_{\cdot\cdot} - \bar{\boldsymbol{\mu}}) - \hat{\mathbf{C}}\mathfrak{S}_{\cdot\cdot}\right| = o_P(1), \tag{4.9}$$

where

$$\begin{aligned}
\mathfrak{S}_{\cdot\cdot} &= \left(\sum_{i=1}^{k} N_i \boldsymbol{\Sigma}_i^{-1}\right)^{-1} \sum_{i=1}^{k} N_i \boldsymbol{\Sigma}_i^{-1} N_i^{-1/2} \mathfrak{S}_{i,N_i} \\
&= \hat{\mathbf{C}}^{-1} \left(\sum_{i=1}^{k} N_i \hat{\mathbf{C}} \boldsymbol{\Sigma}_i^{-1} \hat{\mathbf{C}}\right)^{-1} \sum_{i=1}^{k} N_i \hat{\mathbf{C}} \boldsymbol{\Sigma}_i^{-1} \hat{\mathbf{C}} N_i^{-1/2} \hat{\mathbf{C}} \mathfrak{S}_{i,N_i}.
\end{aligned}$$

Combining (4.7)–(4.9), we conclude

$$\hat{T}_N = \sum_{i=1}^{k} N_i \left(N_i^{-1/2}\mathfrak{S}_{i,N_i} - \mathfrak{S}_{\cdot\cdot}\right)^T \boldsymbol{\Sigma}_i^{-1} \left(N_i^{-1/2}\mathfrak{S}_{i,N_i} - \mathfrak{S}_{\cdot\cdot}\right) + o_P(1),$$

and

$$N_i^{1/2}\mathfrak{S}_{\cdot\cdot} = \left(\frac{N_i}{N}\right)^{1/2} \left(\sum_{\ell=1}^{k} \frac{N_\ell}{N} \boldsymbol{\Sigma}_i^{-1}\right)^{-1} \sum_{\ell=1}^{k} \left(\frac{N_\ell}{N}\right)^{1/2} \boldsymbol{\Sigma}_i^{-1} \mathfrak{S}_{i,N_i}.$$

Now (4.19) follows from (4.5) and Lemma 4.8.1. The result in (4.19) can be established in the same way, only (4.4) needs to be replaced with (4.3).

4.7 Proofs of the Theorems 4.3.1 and 4.3.2

Proof of Theorem 4.3.1. Theorem 2 of Horváth, Kokoszka, and Reeder (2013) yields that under H_A, we have that

$$\|\hat{D}_{N,p} - D\| = o_P(1). \tag{4.1}$$

Now (4.1) implies, along the lines of the arguments used in the proof of Theorem 4.2.1, that even under H_A

$$\max_{1\le i\le d} \|\tilde{\varphi}_i - \hat{c}_i\varphi_i\| \overset{P}{\to} 0 \text{ with } \tilde{c}_i = \operatorname{sign}(\langle\tilde{\varphi}_i, \varphi_i\rangle)$$

and

$$\left|\tilde{\boldsymbol{\Sigma}}_i - \tilde{\mathbf{C}}\boldsymbol{\Sigma}\tilde{\mathbf{C}}\right| = o_P(1).$$

According to Lemma 4.9.1, the matrices $\boldsymbol{\Sigma}_i$ are nonsingular. So by the ergodic theorem in Hilbert spaces (cf. Horváth, Hušková, and Rice (2013)), we conclude

$$(\tilde{\boldsymbol{\xi}}_{\ell\cdot} - \tilde{\boldsymbol{\xi}}_{\cdot\cdot})^T \tilde{\boldsymbol{\Sigma}}_\ell^{-1} (\tilde{\boldsymbol{\xi}}_{\ell\cdot} - \tilde{\boldsymbol{\xi}}_{\cdot\cdot}) \overset{P}{\to} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\cdot\cdot})^T \boldsymbol{\Sigma}_\ell^{-1} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\cdot\cdot}), \;\; 1 \le \ell \le k, \tag{4.2}$$

where $\boldsymbol{\mu}_\ell = (\langle\mu_\ell, \varphi_1\rangle, \langle\mu_\ell, \varphi_2\rangle, \ldots, \langle\mu_\ell, \varphi_d\rangle)^T, \;\; 1 \le \ell \le k$ and

$$\boldsymbol{\mu}_{\cdot\cdot} = \left(\sum_{i=1}^{k} a_i \boldsymbol{\Sigma}_i^{-1}\right)^{-1} \sum_{i=1}^{k} a_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i.$$

Assumption (4.1) implies that $\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for some $1 \leq i, j \leq k$ and therefore, for some $1 \leq \ell \leq k$, $\boldsymbol{\mu}_\ell \neq \boldsymbol{\mu}_{\cdot\cdot}$. Now the result follows from (4.2). □

Proof of Theorem 4.3.2. Elementary arguments give $\tilde{D}_N(t,s) = D^*_{N,p}(t,s) + Q_{N,1}(t,s) + Q_{N,2}(t,s)$, where

$$D^*_{N,p}(t,s) = \sum_{i=1}^{k} \frac{N_i}{N}\left(\gamma_{i,0}(t,s) + \sum_{\ell=1}^{\infty} K(\ell/h)(\gamma_{i,\ell}(t,s) + \gamma_{i,\ell}(s,t))\right)$$

with $\gamma_{i,\ell}(t,s) = \frac{1}{N_i}\sum_{j=1}^{N_i-\ell}(X_{i,j}(t) - \mu_i(t))(X_{i,j+\ell}(s) - \mu_i(s))$, $Q_{N,1}(t,s) = Q_{N,1,1}(t,s) + Q_{N,1,1}(s,t)$ with

$$Q_{N,1,1}(t,s) = \sum_{i=1}^{k} \frac{N_i}{N}(\mu_i(t) - \bar{X}_{\cdot\cdot}(t)) \sum_{\ell=1}^{\infty} K(\ell/h) \frac{1}{N_i} \sum_{j=1}^{N_i-\ell}(X_{i,j+\ell}(s) - \mu_i(s))$$

and

$$Q_{N,2}(t,s) = 2\sum_{\ell=1}^{\infty} K(\ell/h) \sum_{i=1}^{k}(\mu_i(t) - \bar{X}_{\cdot\cdot}(t))(\mu_i(s) - \bar{X}_{\cdot\cdot}(s)).$$

It follows from Horváth, Kokoszka, and Reeder (2013) that

$$\|D^*_{N,p} - D\| = o_P(1). \tag{4.3}$$

By the ergodic theorem in Hilbert spaces, (cf. Horváth, Hušková, and Rice (2013)), we get that $\max_{1\leq i\leq k}\|\mu_i - \bar{X}_{\cdot\cdot}\| = O_P(1)$, and therefore,

$$\|Q_{N,1,1}\| = O_P(1) \max_{1\leq i\leq k} \sum_{\ell=1}^{\infty} K(\ell/h) \left\| \frac{1}{N_i} \sum_{j=1}^{N_i-\ell}(X_{i,j+\ell}(s) - \mu_i(s)) \right\|.$$

Assumptions (4.1)–(4.4) imply (cf. Hörmann and Kokoszka (2010)) that

$$\max_{1\leq i\leq k} \max_{1\leq \ell\leq ch} E\left\| \frac{1}{N_i} \sum_{j=1}^{N_i-\ell}(X_{i,j+\ell}(s) - \mu_i(s)) \right\|^2 = O\left(\frac{1}{N}\right),$$

resulting in $\|Q_{N,1,1}\| = O_P(h/N^{1/2})$. Thus we get

$$\|Q_{N,1}\| = O_P(h/N^{1/2}) = o_P(1). \tag{4.4}$$

Using again the central limit theorem in Hilbert spaces, we have that $\|\bar{X}_{\cdot\cdot} - \bar{\mu}\| = O_P(N^{-1/2})$ and therefore, we conclude

$$\left\| Q_{N,2}(t,s) - 2\sum_{\ell=1}^{\infty} K(\ell/h) \sum_{i=1}^{k} M_i(t)M_i(s) \right\| = O_P(hN^{-1/2}) = o_P(1), \tag{4.5}$$

where $M_i(t)$ is defined (4.3). Combining (4.3)–(4.5), we conclude

$$\left\| \tilde{D}_N(t,s) - \left(2\sum_{\ell=1}^{\infty} K(\ell/h)\right) \sum_{i=1}^{k} M_i(t)M_i(s) + D(t,s) \right\| = o_P(1). \tag{4.6}$$

By using Lemma 4.9.3 with $M = 2\sum_{\ell=1}^{\infty} K(\ell/h)$, we obtain that

$$\max_{1\le i\le m} \|\hat{\varphi}_i - \hat{c}_i\psi_i\| = O_P(1/h) \tag{4.7}$$

and

$$\max_{m+1\le i\le d} \|\hat{\varphi}_i - \hat{c}_i g_{i-m}\| = o_P(1) \tag{4.8}$$

which imply that

$$\max_{1\le i\le k} |\hat{\mathbf{\Sigma}}_i - \widehat{\mathbf{C}}\bar{\mathbf{H}}_i\widehat{\mathbf{C}}| = o_P(1), \tag{4.9}$$

where $\widehat{\mathbf{C}} = \operatorname{diag}(\hat{c}_1, \ldots, \hat{c}_d)$ and $\bar{\mathbf{H}}_i = \left\{ \iint D_i(t,s)\varphi_\ell^*(t)\varphi_k^*(s)dtds,\ 1 \le \ell, k \le d \right\}$ with $\varphi_i^* = \psi_i,\ 1 \le i \le m$ and $\varphi_i^* = g_{i-m},\ m+1 \le i \le d$. By Lemma 4.9.1, the matrices $\bar{\mathbf{H}}_i$ are nonsingular. By the law of large numbers in Hilbert spaces, (4.7)–(4.9) we get that $|\bar{\boldsymbol{\xi}}_{i\cdot} - \bar{\boldsymbol{\xi}}_{\cdot\cdot} - \widehat{\mathbf{C}}(\boldsymbol{\mu}_i^* - \boldsymbol{\mu}_{\cdot\cdot}^*)| = o_P(1)$, where $\boldsymbol{\mu}_\ell^* = (\langle \mu_\ell, \varphi_1^*\rangle, \langle \mu_\ell, \varphi_2^*\rangle, \ldots, \langle \mu_\ell, \varphi_d^*\rangle)^T,\ \ 1 \le \ell \le k$ and

$$\boldsymbol{\mu}_{\cdot\cdot}^* = \left(\sum_{i=1}^{k} a_i \mathbf{H}_i^{-1}\right)^{-1} \sum_{i=1}^{k} a_i \mathbf{H}_i^{-1}\boldsymbol{\mu}_\ell^*.$$

We write $\boldsymbol{\mu}_i^* - \boldsymbol{\mu}_{\cdot\cdot}^* = \bar{\boldsymbol{\mu}}_i^* + \mathbf{c}$, where $\bar{\boldsymbol{\mu}}_\ell^* = (\langle \mu_\ell - \bar{\mu}, \varphi_1^*\rangle, \langle \mu_\ell - \bar{\mu}, \varphi_2^*\rangle, \ldots, \langle \mu_\ell - \bar{\mu}, \varphi_d^*\rangle)^T,\ 1 \le \ell \le k$ and $\mathbf{c} = (\langle \bar{\mu}, \varphi_1^*\rangle, \langle \bar{\mu}, \varphi_2^*\rangle, \ldots, \langle \bar{\mu}, \varphi_d^*\rangle)^T - \boldsymbol{\mu}_{\cdot\cdot}^*$. If $\bar{\boldsymbol{\mu}}_\ell^* + \mathbf{c} \ne \mathbf{0}$ for some $1 \le \ell \le k$, the result follows since $\mathbf{H}_i$ is (strictly) positive definite for all $1 \le i \le k$. If $\bar{\boldsymbol{\mu}}_\ell^* + \mathbf{c} = \mathbf{0}$ for all $1 \le \ell \le k$, then $\bar{\boldsymbol{\mu}}_1^* = \bar{\boldsymbol{\mu}}_2^* = \ldots = \bar{\boldsymbol{\mu}}_k^*$. This yields that $\mathbf{c} = \mathbf{0}$ and therefore, $\bar{\boldsymbol{\mu}}_\ell^* = \mathbf{0}$ for all $1 \le \ell \le k$. However, (4.5) with Lemma 4.9.2, there is i such that $\langle M_i, \varphi_1^*\rangle \ne 0$ and therefore, $\bar{\boldsymbol{\mu}}_\ell^* \ne \mathbf{0}$ for at least one ℓ. □

4.8 Distribution of a quadratic form of normal vectors

Suppose in this section that $\mathbf{Z}_1, ..., \mathbf{Z}_k$ are independent normal random vectors in $\mathbb{R}^d$ so that $E\mathbf{Z}_i = \mathbf{0}$ for all $1 \le i \le$, and $E\mathbf{Z}_i\mathbf{Z}_i^T = \mathbf{\Sigma}_i,\ \ 1 \le i \le k$. Define

$$\boldsymbol{\zeta} = \left(\sum_{\ell=1}^{k} c_\ell \mathbf{\Sigma}_\ell^{-1}\right)^{-1} \sum_{\ell=1}^{k} c_\ell^{1/2}\mathbf{\Sigma}_\ell^{-1}\mathbf{Z}_\ell,$$

where $c_i, 1 \le i \le k$ satisfy $\sum_{i=1}^{k} c_i = 1$, and $c_i > 0$ for all $1 \le i \le k$. We recall that $\chi^2(r)$ stands for a χ^2 random variable with r degrees of freedom.

Lemma 4.8.1. *If* $T = \sum_{\ell=1}^{k} (\mathbf{Z}_l - c_\ell^{1/2}\boldsymbol{\zeta})^T \boldsymbol{\Sigma}_\ell^{-1} (\mathbf{Z}_\ell - c_\ell^{1/2}\boldsymbol{\zeta})$, *then we have that* $T \overset{\mathcal{D}}{=} \chi^2(d(k-1))$.

Proof. The result follows from Theorem 2.7 in Seber and Lee (2003, p. 28) (cf. also Gamage et al.. (2004) and Krishnamoorthy and Lu (2010)).

□

4.9 Three technical lemmas

Lemma 4.9.1. *Let* $Y(t) \in L^2$ *with* $EY(t) = 0$, $E\|Y\|^2 < \infty$ *and* $H(t,s) = EY(t)Y(s)$ *be a strictly positive function and* $\{\kappa_i, 1 \le i < \infty\}$ *be orthonormal functions. Then for any* $1 \le d < \infty$, *the matrix* $\mathbf{C} = \{E\langle Y, \kappa_i\rangle\langle Y, \kappa_j\rangle, 1 \le i, j \le d\}$ *is nonsingular.*

Proof. The matrix $\mathbf{C}$ is the covariance matrix of the random vector $(\langle Y, \kappa_1\rangle, \langle Y, \kappa_2\rangle, \ldots, \langle Y, \kappa_d\rangle)^T$. If $\mathbf{C}$ is singular, then one of the coordinates is a linear combination of the others so that $\sum_{i=1}^{d} d_i\langle Y, \kappa_i\rangle = 0$ where at least one of the d_is is different from 0. Now with $g(t) = \sum_{i=1}^{d} d_i g_i(t)$ we can write

$$0 = \operatorname{var}\left(\sum_{i=1}^{d} d_i\langle Y, \kappa_i\rangle\right) = \operatorname{var}(\langle Y, g\rangle) = \iint H(t,s)g(t)g(s)dtds. \tag{4.1}$$

But $g(t)$ is a linear combination of orthogonal functions where some of the coefficients differ from 0, and therefore, g is not the zero function in L^2. Hence (4.1) contradicts the assumption that H is strictly positive definite. □

Lemma 4.9.2. *We assume* $m \ge 1$, $g_1, g_2, \ldots, g_m \in L^2$, b_1, b_2, b_m *are non–negative numbers and* $U(t,s)$ *is a symmetric positive definite function with eigenvalues* $\gamma_1 > \gamma_2 > \ldots > \gamma_\ell > \gamma_{\ell+1} \ge \ldots 0$ *and corresponding orthonormal eigenfunctions* $\kappa_1, \kappa_2, \ldots$ *Let* $U^*(t,s) = \sum_{i=1}^{m} b_i g_i(t)g_i(s) + U(t,s)$, *with eigenvalues* $\gamma_1^* \ge \gamma_2^* \ge \ldots \ge 0$ *and corresponding orthonormal eigenfunctions* $\kappa_1^*, \kappa_2^*, \ldots$. *If* $\max_{1\le i\le m} b_i\|g_i\|^2 > \lambda_\ell$, *then with some* $j = 1, 2, \ldots, \ell$ *and* $i = 1, 2, \ldots, m$ *we have that*

$$\langle \kappa_j^*, g_i\rangle \ne 0. \tag{4.2}$$

Proof. Suppose by way of contradiction that $\langle \kappa_j^*, g_i \rangle = 0$ for all $1 \le i \le m,\ 1 \le j \le \ell$. This assumption yields by simple calculations that $\kappa_1^*, \dots, \kappa_\ell^*$ are also eigenfunctions of U with eigenvalues $\gamma_1^* \ge \gamma_2^* \ge \dots \ge \gamma_\ell^*$ and

$$\gamma_j^* = \iint U^*(t,s)\kappa_j^*(t)\kappa_j^*(s)dtds = \iint U(t,s)\kappa_j^*(t)\kappa_j^*(s)dtds \text{ for all } 1 \le j \le \ell.$$

Also,

$$\begin{aligned}
\gamma_1 &= \iint U(t,s)\kappa_1(t)\kappa_1(s)dtds \\
&\le \iint U(t,s)\kappa_1(t)\kappa_1(s)dtds + \iint \left[\sum_{i=1}^m b_i g_i(t) g_i(s)\right] \kappa_1(t)\kappa_1(s)dtds \\
&= \iint U^*(t,s)\kappa_1(t)\kappa_1(s)dtds \\
&\le \sup_{\|\kappa\|=1} \iint U^*(t,s)\kappa(t)\kappa(s)dtds \\
&= \gamma_1^*.
\end{aligned}$$

Since $\gamma_1,\ \gamma_1^*$ are both eigenvalues of U and γ_1 is the unique largest, we conclude that $\gamma_1 = \gamma_1^*$. By the uniqueness of γ_1, we also get that $\kappa_1(t) = \pm\kappa_1^*(t)$. Let $\mathcal{S}_j = \text{span}(\kappa_i, 1 \le i \le j)$, $\bar{\mathcal{S}}_j = \{\kappa :\ \|\kappa\| = 1, \langle \kappa, \zeta \rangle = 0 \text{ for all } \zeta \in \mathcal{S}_j\}$ and $\mathcal{S}_j^* = \text{span}(\kappa_i^*, 1 \le i \le j)$, $\bar{\mathcal{S}}_j^* = \{\kappa :\ \|\kappa\| = 1, \langle \kappa, \zeta \rangle = 0 \text{ for all } \zeta \in \mathcal{S}_j^*\}$. Clearly, $\bar{\mathcal{S}}_1 = \bar{\mathcal{S}}_1^*$ and therefore, $\kappa_2 \in \bar{\mathcal{S}}_1^*$. So by Debnath and Mikusiński (2005, p. 197), we get

$$\begin{aligned}
\gamma_2 &= \iint U(t,s)\kappa_2(t)\kappa_2(s)dtds \\
&\le \iint U(t,s)\kappa_2(t)\kappa_2(s)dtds + \iint \left[\sum_{i=1}^m b_i g_i(t) g_i(s)\right] \kappa_2(t)\kappa_2(s)dtds \\
&= \iint U^*(t,s)\varphi_2(t)\varphi_2(s)dtds \\
&\le \sup_{\kappa \in \bar{\mathcal{S}}_1^*} \iint U^*(t,s)\kappa(t)\kappa(s)dtds \\
&= \lambda_2^*.
\end{aligned}$$

Now we conclude that $\gamma_2 = \gamma_2^*$ and $\kappa_2(t) = \pm\kappa_2^*(t)$. Repeating the arguments above, one can easily verify that $\gamma_j = \gamma_j^*$ and $\kappa_j(t) = \pm\kappa_j^*(t)$ for all $1 \le j \le \ell$. Also, since $g_i \in \bar{\mathcal{S}}_{\ell-1}^* = \bar{\mathcal{S}}_{\ell-1}$, we get that for all $1 \le i \le m$

$$\gamma_\ell = \gamma_\ell^* = \sup_{\kappa \in \bar{\mathcal{S}}_{\ell-1}} \iint U^*(t,s)\kappa(t)\kappa(s)dtds$$
$$\geq \iint U^*(t,s)\left[g_i(t)g_i(s)/\|g_i\|^2\right]dtds$$
$$\geq b_i\|g_i\|^2,$$

which contradicts (4.2). □

We assume that

$$\kappa_1, \kappa_2, \ldots, \kappa_m \text{ are orthonormal functions} \tag{4.3}$$

and

$$e_1 > e_2 > \ldots > e_m > 0. \tag{4.4}$$

Let $\mathcal{A}_0 = \text{span}(\kappa_1, \kappa_2, \ldots, \kappa_m)$. We recall that $\bar{\mathcal{B}}$ denotes the orthogonal complement of the set $\mathcal{B}$. Assume that

$$D \text{ is symmetric, square integrable on } [0,1]^2, \text{ and non-negative definite.} \tag{4.5}$$

We say that D has regular maxima of order n with respect to $\mathcal{A}_0$ if there are $r_1 > r_2 > \ldots > r_n$ and orthonormal function $g_1, g_2, \ldots, g_n$ such that

$$r_i = \sup_{g \in \bar{\mathcal{A}}_{i-1}: \|g\|=1} \iint g(t)D(t,s)g(s)dtds = \iint g_i(t)D(t,s)g_i(s)dtds, \ \ 1 \leq i \leq n,$$

with $\mathcal{A}_i = \text{span}(\kappa_1, \ldots, \kappa_m, g_1, \ldots, g_i), 1 \leq i \leq n-1$. The functions $g_1, \ldots, g_n$ are unique up to signs.

Let

$$D_M(t,s) = M\sum_{i=1}^m e_i\kappa_i(t)\kappa_i(s) + D(t,s), \ \ 0 \leq t, s \leq 1.$$

Since D_M is symmetric, non-negative definite, there are $\lambda_{1,M} \geq \lambda_{2,M} \geq \ldots \geq 0$ and orthonormal functions $f_{1,M}, f_{2,M}, \ldots$ such that $\lambda_{i,M}f_{i,M}(t) = \int D_M(t,s)f_{i,M}(s)ds$.

Lemma 4.9.3. *If* (4.3)–(4.5) *hold and D has regular maxima of order n with respect to $\mathcal{A}_0$, then, as $M \to \infty$ we have*

$$\max_{1\leq i\leq m}\|f_{i,M} - c_i\kappa_i\| = o(1), \tag{4.6}$$

$$\max_{1\leq i\leq m}|\lambda_{i,M}/M - e_i| = o(1) \tag{4.7}$$

and

$$\max_{m<i\leq m+n}\|f_{i,M} - c_i g_{i-m}\| = o(1), \tag{4.8}$$

$$\max_{m<i\leq n}|\lambda_{i,M}-e_{i-m}|=o(1), \tag{4.9}$$

where the values of $c_1=c_{1,M}, c_2=c_{2,M},\ldots,c_{m+n}=c_{m+n,M}$ *are* 1 *or* -1.

Proof. Clearly,

$$\left\|\frac{1}{M}D_M(t,s)-\sum_{i=1}^{m}e_i\kappa_i(t)\kappa_i(s)\right\|=O\left(\frac{1}{M}\right),\ \text{ as }\ M\to\infty$$

and therefore, (4.6) and (4.7) follow from Debnath and Mikusiński (2005). Let $\mathcal{A}_{0,M}=\operatorname{span}(f_{1,M},f_{2,M},\ldots,f_{m,M})$. Also,

$$\sup_{f\in\bar{\mathcal{A}}_{0,M},\|f\|=1}\iint D_M(t,s)f(t)f(s)dtds=\iint D_M(t,s)f_{m+1,M}(t)f_{m+1,M}(s)dtds.$$

Every $f\in\bar{\mathcal{A}}_{0,M}$ can be written as

$$f(t)=\omega(t)+f^*(t),\ \text{ where }\ \omega(t)=\sum_{i=1}^{m}\alpha_i\kappa_i(t)\ \text{ and }\ f^*\in\bar{\mathcal{A}}_0.$$

It follows from the definition of D_M that

$$\begin{aligned}&\iint D_M(t,s)f(t)f(s)dtds-\iint D(t,s)f^*(t)f^*(s)dtds\\&\quad=\iint D_M(t,s)\omega(t)\omega(s)dtds+2\iint D_M(t,s)\omega(t)f^*(s)dtds.\end{aligned}$$

Since $\|f^*\|\leq 1$, by the Cauchy–Schwarz inequality, we have

$$\left|\iint D_M(t,s)\omega(t)f^*(s)dtds\right|=\left|\iint \omega(t)D(t,s)f^*(s)dtds\right|=O(1)\|\omega\|,$$

we conclude

$$\left|\iint D_M(t,s)f(t)f(s)dtds-\iint D(t,s)f^*(t)f^*(s)dtds\right|=O(1)(M\|\omega\|^2+\|\omega\|). \tag{4.10}$$

We note that $\langle\omega,\omega\rangle=\sum_{i=1}^m\alpha_i^2$. Since g is orthogonal for all $f_{i,M}, 1\leq i\leq m$, for each $1\leq i\leq m$, we get

$$\begin{aligned}0&=\langle\sum_{\ell=1}^{m}\alpha_\ell\kappa_\ell+f^*,f_{i,M}\rangle\\&=\langle\sum_{\ell=1}^{m}\alpha_\ell\kappa_\ell+f^*,f_{i,M}-c_i\kappa_i\rangle+\langle\sum_{\ell=1}^{m}\alpha_\ell\kappa_\ell+f^*,c_i\kappa_i\rangle\\&=\langle\sum_{\ell=1}^{m}\alpha_\ell\kappa_\ell+f^*,f_{i,M}-c_i\kappa_i\rangle+c_i\alpha_i\end{aligned}$$

resulting in

$$|\alpha_i| \leq \left|\langle \sum_{\ell=1}^{m} \alpha_\ell \kappa_\ell + f^*, f_{i,M} - c_i \kappa_i \rangle \right| \leq \|f_{i,M} - c_i \kappa_i\| = O(1/M).$$

Thus we get

$$\|\omega\|^2 = O(1/M^2)$$

and therefore, by (4.10)

$$\left| \iint D_M(t,s) f(t) f(s) dt ds - \iint D(t,s) f^*(t) f^*(s) dt ds \right| = O\left(\frac{1}{M}\right). \tag{4.11}$$

This means that

$$\left\| \sup_{f \in \bar{\mathcal{A}}_{0,M}, \|f\|=1} \iint D_M(t,s) f(t) f(s) dt ds - r_1 \right\| = \left(\frac{1}{M}\right),$$

which yields $|\lambda_{m+1,M} - e_1| = o(1)$. We assumed that r_1 is reached at only at $\pm g_1$ and thus we conclude that

$$\|f_{m+1,M} - c_{m+1,M} g_1\| = o(1), \quad \text{as} \quad M \to \infty.$$

Any $f \in \bar{\mathcal{A}}_{1,M}$ can be written as

$$f(t) = \sum_{\ell=1}^{m} \beta_\ell \kappa_\ell + \beta_{m+1} g_1(t) + f_1^*, \quad \text{where} \quad f_1^* \in \bar{\mathcal{A}}_1.$$

Following the arguments leading to (4.11)

$$\begin{aligned} \Bigg| \iint & D_M(t,s) f(t) f(s) dt ds \\ & - \iint D(t,s)(\beta_{m+1,M} g_1(t) + f_1^*(t)(\beta_{m+1,M} g_1(s) + f_1^*(s)) dt ds \Bigg| \\ = O & \left(\frac{1}{M}\right). \end{aligned} \tag{4.12}$$

Since $f \in \bar{\mathcal{A}}_{1,M}$, we obtain that $|\beta_{m+1,M}| = O(1)\|f_{m+1,M} - c_{m+1,M} g_1\| = o(1)$. Since r_2 is reached uniquely at $\pm g_2$, we conclude that $|\lambda_{m+2,M} - e_2| = o(1)$ and $\|f_{m+2,M} - c_{m+2,M} g_2\| = o(1)$, as $M \to \infty$. Repeating the arguments above, (4.8) and (4.9) of Theorem 4.9.3 can be established for $i = m+3, \ldots, m+n$. □

4.10 Bibliography

[1] Abramovich, F. and Angelini, C.: Testing in mixed–effects FANOVA models. *Journal of Statistical Planning and Inference* 136(2006), 4326–4348.

[2] Andrews, D.W.K.: Heteroskedasticity and autocorellation consistent covariance matrix estimation. *Econometrica* 59 (1991), 817–858.

[3] Andrews, D.W.K. and Monahan, J.C.: An improved heteroskedasticity and autocorellation consistent covariance matrix estimatior. *Econometrica* 60 (1992), 953–966.

[4] Antoniadis, A. and Sapatinas, T.: Estimation and inference in functional mixed–effect models. *Computational Statistics and Data Analysis* 51(2007), 4793–4813.

[5] Aue, A., Hörmann, S., Horváth, L., Hušková, M. and J. Steinebach, J.: Sequential testing for the stability of portfolio betas. *Econometric Theory* 28(2012), 804–837.

[6] Berkes, I., Horváth, L. and Rice, G.: Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications* 123(2013), 385–403.

[7] Bosq, D.: *Linear Processes in Function Spaces.* Springer Verlag, 2000.

[8] Cuevas, A., Febrero, M. and Freiman, R.: An anova test for functional data. *Computational Statistics and Data Analysis* 47(2004), 111–122.

[9] Debnath, L. and Mikusiński, P.: *Hilbert Spaces with Applications.* Third Edition, Elsevier, New York 2005.

[10] Doukhan, P., Massart, P. and Rio, E.: Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincare B* 31(1995), 393–427.

[11] Drignei, D.: Functional ANOVA in computer models with time series output. *Technometrics* 52, 430–437.

[12] Gamage, J. Mathew, T. and Weerahandi, S.: Generalized p–values and generalized confidence regoins for the multivariate Behrens–Fisher problem and MANOVA. *Journal of Multivariate Analysis* 88(2004), 177–189.

[13] Gohberg, I., Golberg, S. and Kaashoek, M.A.: *Classes of Linear Operators.* Operator Theory: Advances and Applications. vol. 49, Birkhäuser, 1990.

[14] Hall, P. and Keilegom, Van I.: Two–sample tests in functional data analysis starting from discrete data. *Statistica Sinica* 17(2007), 1511–1531.

[15] Hannan, E.J.: Central limit theorems for time series regression. *Z. Wahrsch. verw. Gebiete* 26(1973), 157–170.

[16] Hooker, G.: Generalized functional ANOVA diagnostics for high–dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics* 16(2007), 709–732.

[17] Hörmann, S. and Kokoszka, P.: Weakly dependent functional data. *Annals of Statistics* 38 (2010), 1845–1884.

[18] Hörmann, S., Horváth, L. and Reeder, R.: A functional version of the ARCH model. *Econometric Theory* 29(2013), 138–152.

[19] Horváth, L., Hušková and Rice, G.: Test of independence for functional data. *Journal of Multivariate Analysis* 117(2013), 100-119.

[20] Horváth, L. and Kokoszka, P.: *Inference for Functional Data with Applications.* Springer, New York, 2012.

[21] Horváth, L. Kokoszka, P. and Reeder, R.: Estimation of the mean of of functional time series and a two sample problem. *Journal of the Royal Statistical Society Ser. B* 75(2013), 103–122.

[22] Horváth, L., Kokoszka, P. and Reimherr, M.: Two sample inference in functional linear models. *Canadian Journal of Statistics* 37(2009), 571–591.

[23] Jirak, M.: On weak invariance principles for sums of dependent random functionals. *Statistics & Probability Letters* 83(2013), 2291–2296.

[24] Krishnamoorthy, K. and Lu, F.: A parametric bootstrap solution to the MANOVA under heteroscedasticity. *Journal of Statistical Computation and Simulation* 80(2009), 873–887.

[25] Laukaitis, A. and Račkauskas, A.: Functional data analysis for client segmentation tasks. *European Journal of Operational Research* 163(2005), 210–216.

[26] Magnano, L. and Boland, J.W.: Generation of synthetic sequences of electricity demand: Application in South Australia. *Energy* 32(2007), 2230–2243.

[27] Magnano, L., Boland, J.W. and Hyndman, R.J.: Generation of synthetic sequences of half–hourly temperature. *Environmetrics* 19(2008), 818–835.

[28] Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L.: Statistical significance testing for the association of magnetometer records at high-, mid- and low latitudes during substorm days. *Planetary and Space Science* 58(2010), 437-445.

[29] Martínez–Camblor, P. and Corral, N.: Repeated measures analysis for functional data. *Computational Statistics and Data Analysis* 55(2011), 3244-3256.

[30] Politis, D.N.: Adaptive bandwidth choice. *Journal of Nonparametric Statistics* 25(2003), 517–533.

[31] Pötscher, B.M and Prucha, I.R.: Basic structure of the asymptotic theory in dynamic nonlinear econometrics models, Part I: Consistency and approximation concepts. *Econometric Reviews* 10(1991), 125-216.

[32] Ramsay, J.O., Hooker, G. and Graves, S.: *Functional Data Analysis with R and MATLAB(Use R!).* Springer, New York 2009.

[33] Ramsay, J.O. and Silvermann, B.W.: *Functional Data Analysis.* Springer, New York 2005.

[34] R Development Core Team: *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2008, URL `http://www.R-project.org`.

[35] Seber, G.A.F. and Lee, A.J.: *Linear Regression Analysis* (Second Edition). Wiley, New York 2003.

[36] Shang, H. L. and Hyndman, R. J.: *fds: Functional data sets* R package version 2.10.0, 2013.

[37] Wu, W.B. and Min, M.: (2005). On linear processes with dependent innovations. *Stochastic Processes and their Applications* 115(2005), 939-958.

[38] Zhang, J.-T.: *Analysis of Variance for Functional Data.* Chapman & Hall/CRC, 2013.

Table 4.1. Empirical sizes in the i.i.d. case and the FAR cases with nominal levels of 10%, 5%, and 1%.

DGP		IID			FAR_{ψ_1}			FAR_{ψ_2}		
N	Nominal	$k=3$	$k=5$	$k=10$	$k=3$	$k=5$	$k=10$	$k=3$	$k=5$	$k=10$
50	10%	10.5	10.6	10.7	11.2	14.1	17.8	14.4	15.5	18.0
	5%	5.8	5.6	5.7	5.4	7.0	9.6	7.4	9.3	10.1
	1%	1.2	1.3	1.4	.6	1.0	2.2	.6	2.5	2.7
100	10%	10.5	10.1	10.4	12.4	14.0	14.1	11.6	13.2	17.5
	5%	5.5	5.4	5.3	5.8	7.2	7.8	6.7	7.3	9.6
	1%	1.2	1.1	1.1	1.4	1.2	2.1	1.2	1.6	2.7
200	10%	10.5	9.7	9.8	10.7	12.6	13.7	11.0	11.5	15.5
	5%	5.2	4.8	4.9	4.1	6.8	6.5	5.7	6.1	8.9
	1%	1.2	1.1	1.3	1.3	.8	1.7	.9	1.4	2.1
300	10%	12.2	10.4	9.2	9.5	11.5	12.1	10.9	11.1	13.8
	5%	6.5	5.7	4.6	4.4	5.3	5.9	5.6	5.5	7.6
	1%	1.7	.8	1.0	.7	1.4	1.6	1.2	1.4	1.9

Table 4.2. Empirical sizes in with heterogeneous population covariances with nominal levels of 10%, 5%, and 1% in both the independent and dependent cases.

N	50			100			200			300		
DGP/Nominal	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
HIID	9.7	5.7	1.8	11.2	5.9	1.7	11.1	5.4	1.6	10.8	5.3	1.4
HDEP	14.7	6.8	1.3	13.2	5.9	2.1	12.7	5.7	1.6	11.7	5.5	1.7

Table 4.3. Empirical power using i.i.d. and the FAR(1) observations with nominal levels of 10%, 5%, and 1%.

Case		(M1)		(M2)		(M3)		(M4)		(M5)	
N	Level	IID	FAR$_{\psi_2}$	IID	FAR$_{\psi_2}$	IID	FAR$_{\psi_2}$	IID	FAR$_{\psi_2}$	IID	FAR$_{\psi_2}$
25	10%	78.2	65.5	93.0	88.1	82.6	75.2	98.6	88.1	78.6	46.4
	5 %	70.3	54.9	89.5	82.7	73.6	68.0	97.6	83.6	70.0	36.4
	1 %	54.2	36.2	76.9	69.6	55.0	50.1	94.7	68.3	52.8	23.5
50	10%	84.7	69.8	99.0	96.5	90.7	85.2	100	97.4	84.9	51.6
	5 %	78.1	60.3	97.0	92.8	86.6	75.1	100	94.3	78.0	40.3
	1 %	60.5	39.4	92.2	88.0	69.0	58.6	99.9	84.8	60.8	24.7
100	10%	97.1	87.7	99.9	99.9	99.3	96.0	100	99.9	97.8	63.6
	5 %	94.2	79.2	99.7	99.8	98.6	93.5	100	99.8	95.7	54.6
	1 %	85.9	60.6	99.0	99.0	97.0	83.9	100	99.1	88.6	39.2
200	10%	100	99.4	100	100	100	100	100	100	100	77.6
	5 %	100	99.0	100	100	100	100	100	100	100	71.3
	1 %	100	94.6	100	100	100	99.3	100	100	99.7	55.3

Table 4.4. p–values of the FANOVA test when applied to samples of daily LDDCs organized according to the day of the week and season. Across the top of the table, the days included in the sample are displayed. "All" denotes that all seven days were included ($k = 7$).

		Days of the week for which the test is applied				
Season	All	Weekdays	Weekends	TWTh	MTWTh	TWThF
Summer	.000	.035	.006	.750	.647	.056
Fall	.000	.003	.001	.886	.380	.023
Winter	.000	.000	.000	.257	.001	.000
Spring	.000	.002	.000	.582	.083	.001

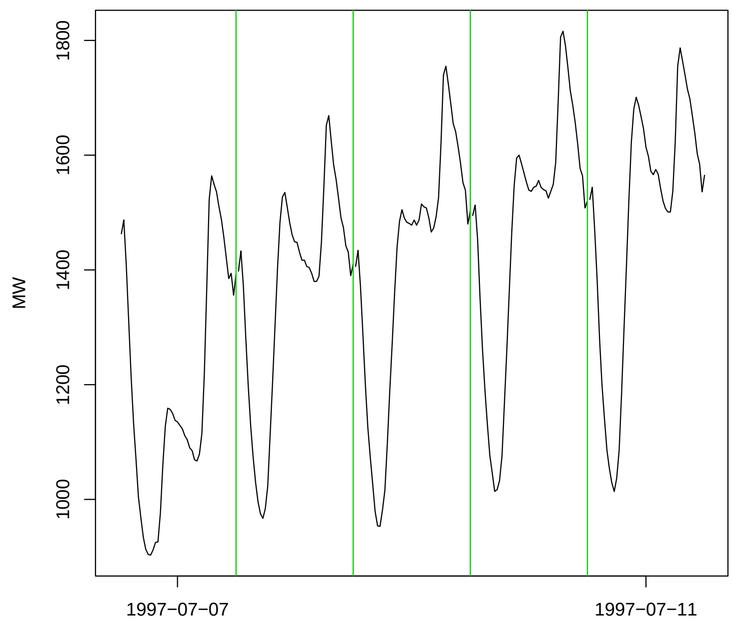

Figure 4.1. Five functional data objects constructed from half-hourly measurements of the electricity demand in Adelaide, Australia. The vertical lines separate the days.

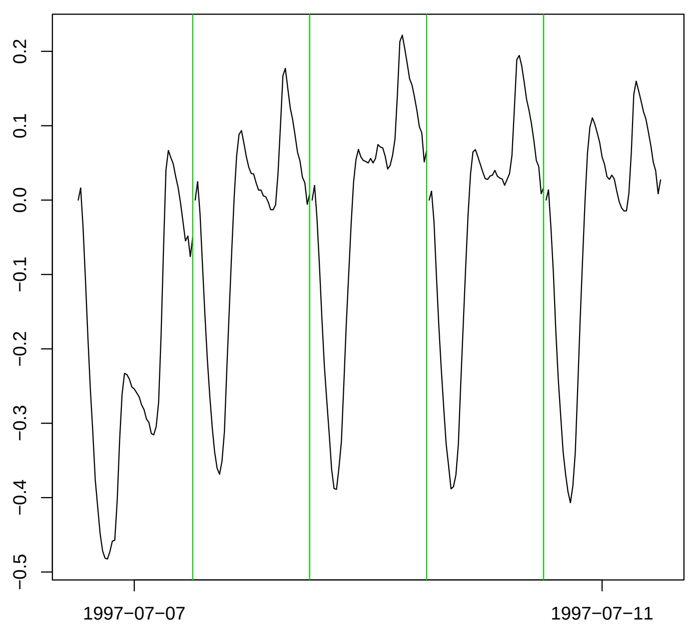

Figure 4.2. Five LDDCs constructed from the curves in Figure 4.1.

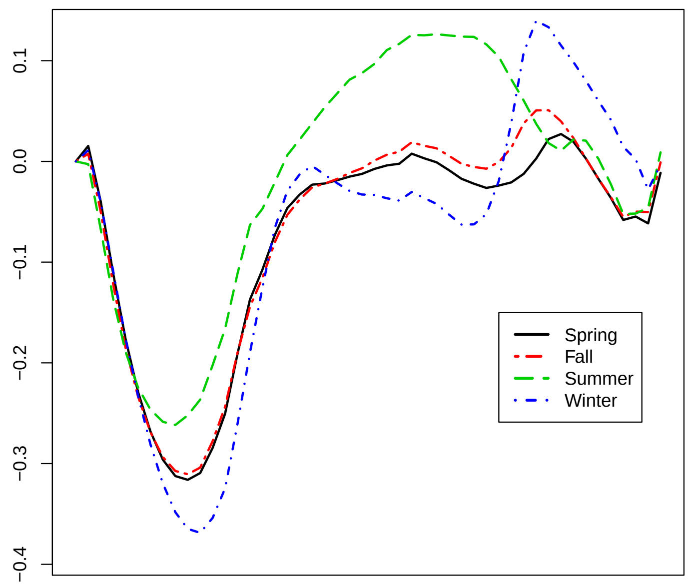

Figure 4.3. Mean Curves from each season constructed from the LDDCs taken from 7/6/1997 to 3/31/2007. The p-value of the FANOVA test applied to this sample was zero.

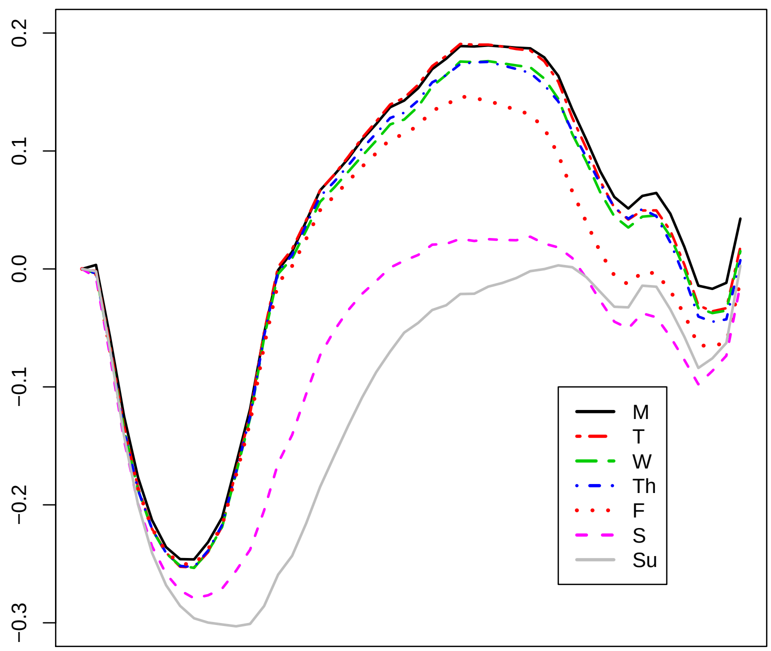

Figure 4.4. Mean curves for each day computed from the Summer months between 7/7/1997 to 7/5/1998 (52 curves for each day). The p-value of the FANOVA test was less than 10^{-4}.